



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification: <b>C12N 15/52, C12N 5/10, C12N 9/00, C12N 15/63, C12P 17/06</b>	<b>A2</b>	(11) International Publication Number: <b>WO 00/22139</b> (43) International Filing Date: 20 April 2000 (20.04.2000)
--	-----------	---

(21) International Application Number: PCT/US99/23535  
(22) International Filing Date: 11 October 1999 (11.10.1999)  
(30) Priority Data:  
198 46 493.2 09 October 1998 (09.10.1998) DE

Published

## (60) Parent Application or Grant

GESELLSCHAFT FÜR BIOTECHNOLOGISCHE  
FORSCHUNG MBH (GBF) [/]; (). BRISTOL-MYERS  
SQUIBB COMPANY [/]; (). BEYER, Stefan [/];  
(). BLOECKER, Helmut [/]; (). BRANDT, Petra [/]; (). CINO,  
Paul, M. [/]; (). DOUGHERTY, Brian, A. [/]; (). GOLDBERG,  
Steven, L. [/]; (). HOFLE, Gerhard [/]; (). MUELLER, Rolf-  
Jochim [/]; (). REICHENBACH, Hans [/]; (). BEYER, Stefan  
[/]; (). BLOECKER, Helmut [/]; (). BRANDT, Petra [/];  
(). CINO, Paul, M. [/]; (). DOUGHERTY, Brian, A. [/];  
(). GOLDBERG, Steven, L. [/]; (). HOFLE, Gerhard [/];  
(). MUELLER, Rolf-Jochim [/]; (). REICHENBACH, Hans  
[/]; (). SANTUCCI, Ronald, R. : ().

(54) Title: DNA SEQUENCES FOR ENZYMATIC SYNTHESIS OF POLYKETIDE OR HETEROPOLYKETIDE COMPOUNDS  
(54) Titre: SEQUENCES D'ADN DESTINEES A LA SYNTHÈSE ENZYMATIQUE DE COMPOSÉS À BASE DE POLYKETIDES  
OU D'HETEROPOLYKETIDES

## (57) Abstract

The invention consists of: (1) cloned *Sorangium cellulosum* polyketide synthase (PKS) biosynthetic cluster DNA; and (2) the nucleotide sequence and predicted protein coding sequences of the cloned DNA. The invention can be used for, but not limited to: (a) increasing yields of PKS product in *Sorangium cellulosum* (e.g., by amplification or genetic modification of the epothilone gene cluster or its component parts); (b) increasing yields of polyketide product in a heterologous system by transfer of the epothilone gene cluster or its component parts, which may be followed by amplification or genetic modification of the PKS gene cluster or its component parts; (c) modification of the polyketide product chemical structure in either *Sorangium cellulosum* or a heterologous host (e.g., by genetic modification of the epothilone gene cluster or its component parts; and (d) for the detection of genes and gene products involved in making polyketides or related molecules in other organisms (e.g., by hybridization or complementation assays). DNA sequence and analysis is presented for the following cosmid and plasmids: A2 cosmid; the pEPOcos6 region (overlapping of pEPOcos6 and pEPOcos7); pEPOcos8 cosmid; A5 cosmid; Sau4 (10 kb plasmid).

## (57) Abrégé

L'invention concerne: (1) un ADN biosynthétique cloné en grappe de polykétide synthase (PKS) de *Sorangium cellulosum*; et (2) la séquence nucléotidique et les séquences protéiques codantes prévues de l'ADN cloné. L'invention peut avoir les applications suivantes (sans caractère limitatif): (a) augmentation de la production de PKS chez *Sorangium cellulosum* (p.ex., par l'amplification ou la modification génétique de la grappe de gènes épothilone ou de ces parties constitutives); (b) augmentation de la production du produit polykétide dans un système hétérologue par le transfert de la grappe de gènes épothilone ou de ces parties constitutives, qui peut être suivie par l'amplification ou la modification génétique de la grappe de gènes PKS ou des ses parties constitutives; (c) modification de la structure chimique du produit polykétide soit chez *Sorangium cellulosum* soit chez un hôte hétérologue (p.ex., par l'amplification ou la modification génétique de la grappe de gènes épothilone ou de ces parties constitutives); et (d) détection de gènes et de produits géniques participant à la fabrication de polykétides ou de molécules correspondantes dans d'autres organismes (p.ex., par des dosages à hybridation ou à complémentation). La séquence d'ADN et l'analyse sont présentées pour les cosmides et les plasmides suivants: cosmide A2; région pEPOcos6 (se chevauchant avec pEPOcos6 et pEPOcos7); cosmide pEPOcos8; cosmide A5; Sau4 (plasmide 10 kb).

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N 15/52, 9/00, 15/63, 5/10, C12P 17/06	A2	(11) International Publication Number: WO 00/22139	(43) International Publication Date: 20 April 2000 (20.04.00)
--	----	---	--

(21) International Application Number: PCT/US99/23535

(22) International Filing Date: 11 October 1999 (11.10.99)

(30) Priority Data:  
198 46 493.2 9 October 1998 (09.10.98) DE

(71) Applicants (for all designated States except US):  
GESELLSCHAFT FÜR BIOTECHNOLOGISCHE  
FORSCHUNG MBH (GBF) [DE/DE]; Mascheroder Weg 1,  
D-38124 Braunschweig (DE). BRISTOL-MYERS  
SQUIBB COMPANY [US/US]; P.O. Box 4000, Princeton,  
NJ 08543-4000 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): BEYER, Stefan [DE/DE];  
Mascheroder Weg 1, D-38124 Braunschweig (DE).  
BLOECKER, Helmut [DE/DE]; Mascheroder Weg 1,  
D-38124 Braunschweig (DE). BRANDT, Petra [DE/DE];  
Mascheroder Weg 1, D-38124 Braunschweig (DE). CINO,  
Paul, M. [US/US]; 4 Crest Drive, Bound Brook, NJ 08805  
(US). DOUGHERTY, Brian, A. [US/US]; 10 Rosemary  
Lane, Killingworth, CT 06419 (US). GOLDBERG, Steven,  
L. [US/US]; 25 Compton Court, Basking Ridge, NJ 07920  
(US). HOFLE, Gerhard [DE/DE]; Mascheroder Weg 1,  
D-38124 Braunschweig (DE). MUELLER, Rolf-Joachim

[DE/DE]; Mascheroder Weg 1, D-38124 Braunschweig  
(DE). REICHENBACH, Hans [DE/DE]; Mascheroder Weg  
1, D-38124 Braunschweig (DE).

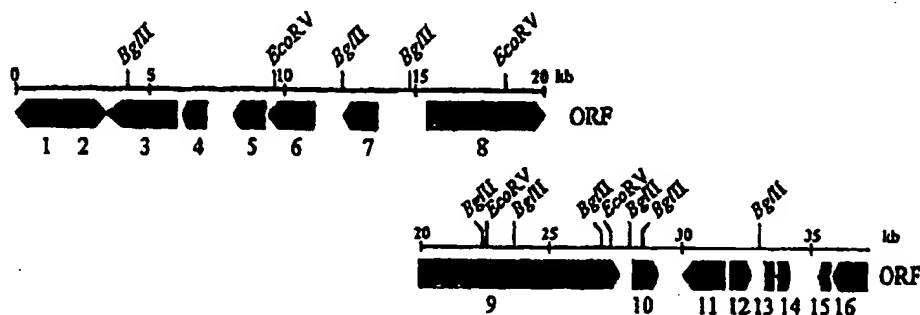
(74) Agents: SANTUCCI, Ronald, R. et al.; Kane, Dalsimer,  
Sullivan and Levy, LLP 20th floor, 711 Third Avenue, New  
York, NY 10017 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG,  
BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE,  
ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP,  
KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA,  
MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU,  
SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG,  
US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE,  
LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM,  
AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT,  
BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU,  
MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM,  
GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

Without international search report and to be republished  
upon receipt of that report.

(54) Title: DNA SEQUENCES FOR ENZYMATIC SYNTHESIS OF POLYKETIDE OR HETEROPOLYKETIDE COMPOUNDS



(57) Abstract

The invention consists of: (1) cloned *Sorangium cellulosum* polyketide synthase (PKS) biosynthetic cluster DNA; and (2) the nucleotide sequence and predicted protein coding sequences of the cloned DNA. The invention can be used for, but not limited to: (a) increasing yields of PKS product in *Sorangium cellulosum* (e.g., by amplification or genetic modification of the epothilone gene cluster or its component parts); (b) increasing yields of polyketide product in a heterologous system by transfer of the epothilone gene cluster or its component parts, which may be followed by amplification or genetic modification of the PKS gene cluster or its component parts; (c) modification of the polyketide product chemical structure in either *Sorangium cellulosum* or a heterologous host (e.g., by genetic modification of the epothilone gene cluster or its component parts; and (d) for the detection of genes and gene products involved in making polyketides or related molecules in other organisms (e.g., by hybridization or complementation assays). DNA sequence and analysis is presented for the following cosmids and plasmids: A2 cosmid; the pEPOcos6 region (overlapping of pEPOcos6 and pEPOcos7); pEPOcos8 cosmid; A5 cosmid; Sau4 (10 kb plasmid).

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LJ	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## D scription

5

10

15

20

25

30

35

40

45

50

55



5

10

15

**DNA sequences for enzymatic synthesis of polyketide or  
heteropolyketide compounds**

10

20

The present invention relates to DNA sequences for enzymatic synthesis of polyketide or heteropolyketide compounds produced by the bacterium *Sorangium cellulosum*.

25

**15 Background and introduction**

30

This patent application describes DNA sequences for the enzymatic synthesis of polyketide and/or heteropolyketide structures synthesized by the myxobacterium *Sorangium cellulosum*. Several of these compounds have known cytotoxic, immunosuppressive, antibiotic and fungicidal biological activity, with the epothilones having been most studied and characterized. The fermentation of large quantities of secondary metabolites from microorganisms, especially from myxobacteria, is a time consuming and difficult process that often involves complications (e.g. contamination, low product yield, difficult isolation and purification). Therefore it would be advantageous to use a well-characterized organism for such fermentations. After cloning of the desired biosynthetic genes one could create such an organism via genetic engineering and manipulate the biosynthesis of the compound. Identified sequences

40

45

30

50

55

5

2

can be cloned into optimized expression vectors and generate recombinant cell lines that overproduce polyketide structures.

10

Polyketide synthases (PKS) and non-ribosomal peptide synthetases (NRPS) represent macromolecular and multifunctional enzymes which are characterized by a modular architecture. PKS condenses activated carbonic acids (usually acetate and propionate) and reduce the resulting 2-keto acid intermediates stepwise in a fatty acid biosynthesis-like fashion. Responsible for each reaction step is a specific domain that recognizes, activates, condenses and reduces the carbonic acid. Depending on the presence of these domains in the corresponding modules, every reduction stage can occur in the final product (Rawlings, Nat. Prod. Reports 14, 523-556 [1997]; for a review, see Chem. Rev. 27, 2463-2760 [1997]). A typical example for the biosynthesis of a polyketide is the macrolide antibiotic erythromycin (Staunton and Wilkinson, Chem. Rev. 27, 2611-2630 [1997]). NRPSs are also modular enzymes and condense via peptide bonds amino acids to low molecular weight bioactive substances like bacitracin or tyrocidin. Typical domains of these systems activate the amino acid and condense it with the growing peptide chain. Methylations, epimerisations and modifications via additional protein domains are possible (Stachelhaus and Marahiel, FEMS Microbiol Lett. 125, 3-14 [1995]). Both types of enzymes (NRPS and PKS) share the modular organization of the proteins in which specific catalytic domains are responsible for recognition, activation, condensation and modification of the single elongation units. The growing chain of amino acids and/or carbonic acids is extended through the action of one module adding one unit. The domains of each module carry the active centers responsible for the enzymatic steps of the biosynthesis.

5

15

20

10

25

15

30

35

40

25

45

50

55

5

3

10

15

20

25

30

35

40

45

50

55

Little is known about the biosynthesis of biologically active polyketides and polypeptides from myxobacteria. Fragments of the biosynthetic gene clusters of soraphen and saframycin have been described from *Sorangium cellulosum* So ce26 and *Myxococcus xanthus*, respectively (Schupp et al., J. Bacteriol. 177, 3673-3679 [1995] and Pospiech et al., Microbiology 141, 1793-1803 [1995]). We have constructed genomic libraries of the epothilone producer *Sorangium cellulosum* So ce90. Gene probes based on PKS and PS genes were used to isolate recombinant cosmids, which were then sequenced and characterized. Several unique pathways containing PKS, PS, or a combination of both types of genes were identified, demonstrating that this organism is potentially a rich source of novel bioactive compounds.

A subject of the present invention is therefore to provide DNA sequences according to claim 1 the expression products of which perform or are involved in the enzymatic biosynthesis, mutasynthesis or partial synthesis of polyketide or heteropolyketide compounds. The DNA sequences may be inserted into well known and optimized expression vectors by common techniques of molecular biology, thus allowing transformation, selection and cloning of cells, which cells are then capable of synthesizing polyketide or heteropolyketide compounds by fermentation. Using an overproducing clone allows the desired polyketide or heteropolyketide compounds be easily produced and recovered in high amounts. Further, knowledge of the localization of regulatory DNA segments and individual structural genes allows "site-directed mutagenesis" using common techniques for genetic engineering, and thus construction of optimized enzymes ("protein engineering") for fermentative synthesis of polyketide or heteropolyketide compounds.

5

4

10

The invention thus further relates to a recombinant expression vector according to claim 16, cells transformed there-  
with according to claim 17 and to a process for enzymatic biosynthesis, mutasynthesis or partial synthesis of polyketide or  
5 heteropolyketide compounds according to claim 23.

15

Preferred and/or advantageous embodiments of the present invention are subject-matter of the subclaims.

20

In brief, the invention consists of (1) cloned *Sorangium*  
*cellulosum* polyketide synthase (PKS) and/or peptide synthetase  
10 (PS) biosynthetic cluster DNA and (2) the nucleotide sequence and predicted protein coding sequences of the cloned DNA. The invention can be used for, but not limited to, (a) increasing  
25 yields of PKS product in *Sorangium cellulosum* (e.g., by amplification or genetic modification of the epothilone gene cluster  
15 or its component parts), (b) increasing yields of polyketide and/or peptide synthetase product in a heterologous system by  
30 transfer of the corresponding gene cluster or its component parts, which may be followed by amplification or genetic modification of the PKS and/or PS gene cluster or its component  
20 parts, (c) modification of the polyketide and/or peptide synthetase product chemical structure in either *Sorangium cellulosum* or a heterologous host (e.g., by genetic modification of  
35 the corresponding gene cluster or its component parts) and (d) for the detection of genes and gene products involved in making  
40 polyketides or related molecules in other organisms (e.g., by hybridization or complementation assays). DNA sequence and  
25 analysis is presented for the following cosmids and plasmids:

45

- A2 cosmid as defined in claim 6
- the pEPOcos6 region (overlapping of pEPOcos6 and pEPOcos7)  
30 as defined in claim 7

50

55

5

5

10

- pEPOcos8 cosmid as defined in claim 10
- A5 cosmid as defined in claim 12
- Sau4 (10 kb plasmid) as defined in claim 14

15

5 The invention is now described in more detail by examples and  
for illustration only. The examples are not to be construed as  
any limitation of the scope.

20

10

Figure 1 is a restriction map of one of the DNA sequences of  
the present invention (cosmid A2 insert) indicating also the  
localization of regulatory DNA segments and the individual  
structural genes ("open reading frames" or ORFs) 1 to 16.

25

15

Figure 2 shows the open reading frames found on pEPOcos6 region

30

DNA sequence data from A2 cosmid are as defined in claim 6.

35

Table 1 correlates ORFs 1 to 16 found on A2 cosmid with the re-  
spective biological function (Regulators, Enzymes).

40

45

50

55

5

6

Table 1

10

15

20

25

30

35

40

45

	gene/function	position
ORF 1	regulatory element	1666 - 1
ORF 2	regulatory element	1605 - 3338
ORF 3	acyl-t-RNA synthetase	6100 - 3398
ORF 4	monooxygenase	7110 - 6374
ORF 5	amino transferase	9590 - 8433
ORF 6	L-dopa decarboxylase	11393 - 9855
ORF 7	oxidoreductase	13656 - 12712
ORF 8	polyketide synthase	15374 - 18984
ORF 9	polypeptide synthetase	20003 - 27889
ORF 10	peptidase	28251 - 29402
ORF 11	regulatory element	31720 - 30401
ORF 12	sigma factor	31982 - 32932
ORF 13	regulatory element	33128 - 33613
ORF 14	regulatory element	33661 - 34007
ORF 15	transcription regulator	35611 - 35255
ORF 16	signal transduction	37856 - 35730

50

55

5

7

## Working Examples

10

A. Construction of a *Sorangium cellulosum* cosmid library5 1. Isolation of genomic DNA from *S. cellulosum* So ce90

15

a. *Sorangium cellulosum* So ce90 was spread onto solid CA-2 agar and incubated at 30°C for 5-7 days. CA-2 agar is prepared by autoclaving 18 g Bacto-agar (Difco Laboratories, Detroit, MI) in 800 ml dH<sub>2</sub>O for 20 min at 121°C and cooling to 50-55°C in a water bath. The following filter-sterilized solutions are added to the agar: 20% (w/v) glucose, 50 ml; Solution A (7.5% [w/v] KNO<sub>3</sub>, 7.5% K<sub>2</sub>HPO<sub>4</sub>), 10 ml; Solution B (1.5% [w/v] MgSO<sub>4</sub>·7H<sub>2</sub>O), 10 ml; Solution C (0.2% [w/v] CaCl<sub>2</sub>·2H<sub>2</sub>O, 0.15% [w/v] FeCl<sub>3</sub>), 10 ml; 1 M HCl, 1 ml; autoclaved 4-day old *Sorangium cellulosum* broth, 100 ml. A sample of cells was removed from the plates with a sterile loop and inoculated into 50 ml of G51t medium in a 250 ml Erlenmeyer flask. G51t consists of 0.5% starch (Cerestar), 0.2% tryptone, 0.1% yeast extract, 0.05% CaCl<sub>2</sub>, 0.05% MgSO<sub>4</sub>·7H<sub>2</sub>O, 1.2% 4-(2-hydroxyethyl)-1-piperazine-ethanesulfonic acid (HEPES), 0.2% glucose, pH 7.6. The flasks were shaken at 30°C, 160 rpm until a dense orange bacterial growth was obtained (ca. 5-7 d.). The cells were pelleted by centrifugation at 6,000 x g and used immediately or stored frozen at -20°C.

30

35

40

The protocol used for isolating chromosomal DNA from bacteria using hexadecyltrimethylammonium bromide (CTAB) has been described previously (Ausubel et al., Current Protocols in Molecular Biology, John Wiley and Sons, New York, 1990). The precipitated DNA was recovered with a bent Pasteur pipette, washed

50

55

5

8

10

with 70% and 95% ethanol, air-dried, and resuspended in 0.5 ml TE buffer (0.01 M Tris-HCl, 0.001 M ethylenediaminetetraacetic acid [EDTA], pH 8.0).

15

20

- 5 b. Alternatively, genomic DNA was isolated from *S. cellulosum* cells cultured as described in section A.1 using the Midi Qiagen Blood & Cell Culture DNA purification Kit (Qiagen, Hilden, Germany) following the Qiagen Genomic DNA Handbook protocol for bacterial DNA isolation (1997, Qiagen, Hilden, Germany, p. 29 ff.). In order to obtain high molecular weight chromosomal DNA the precipitated DNA was recovered with a bent pasteur pipette as described in section A.1.

25

## 15 2. Isolation of plasmid DNA

30

35

40

- a. pFD666: pFD666 is a bifunctional *E. coli*-*Streptomyces* cosmid cloning vector (see Denis and Brzezinski, Gene 111, 115-118 [1992]). To maintain stability of large inserts, it is present in low-medium copy number when replicated in *E. coli*. For this reason, isolation of sufficient pure DNA to carry out cloning experiments was difficult using commercial kits with standard protocols. A modified procedure was therefore used to obtain pFD666 DNA. A 10 ml culture of DH10B(pFD666) was grown for 16-20 hr at 37°C in LB (1% tryptone, 0.5% yeast extract, 0.5% NaCl, pH 7.0) medium containing 50 µg/ml kanamycin sulfate. Fifty ml of LB + kanamycin was inoculated to a starting OD<sub>600</sub> of ca. 0.25 and shaken at 300 rpm, 37°C, until the OD<sub>600</sub> reached ca. 0.6. Five hundred ml of LB + kanamycin medium in a 2 l flask was inoculated with 25 ml of this culture and incubated

45

30

50

55



5

9

10

15

20

25

30

35

40

under the same conditions for 2.5 hr. Chloramphenicol ( 2.5 ml  
of a 34 mg/ml solution in 100% EtOH) was added and the incuba-  
tion continued for an additional 16-20 hr. (The previous steps  
were performed according to Maniatis et al. Molecular Cloning:  
A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring  
Harbor, NY, 1989.) Cells were pelleted for 10 min, 16,000 x g .  
They were resuspended in 9 ml of 50 mM glucose/25 mM Tris-HCl  
(pH 8.0)/10 mM EDTA and transferred to a 50 ml disposable cen-  
trifuge tube. One ml of a freshly-prepared 10 mg/ml lysozyme  
solution in 10 mM Tris-HCl, pH 8.0 was added and the cell sus-  
pension incubated in a 37°C water bath for 10 min. Twenty ml of  
a freshly-prepared 0.2 NaOH/1% sodium dodecyl sulfate (SDS) so-  
lution was added and the tube inverted gently 5-7 times to mix  
the contents. After 5 min at room temperature, 15 ml of 5 M po-  
tassium acetate (pH 4.8) was added and the tube inverted sharply  
3-4 times. The tube was centrifuged at 6,000 x g for 10 min at  
4°C and the supernatant poured through 2 layers of sterile  
cheese cloth into a fresh 50 ml disposable tube. Isopropanol to  
a final concentration of 0.6% was added and the contents of the  
tube mixed several times. The precipitated nucleic acid was  
centrifuged at 6,000 x g for 10 min at 4°C. The pellet was  
washed with 70% EtOH and any excess EtOH was aspirated from the  
pellet, which was allowed to air dry for 5 min. It was resus-  
pended in 5 ml of 50 mM 3-(N-Morpholino)propanesulfonic acid  
(MOPS)/750 mM NaCl, pH 7.0 and added to an equilibrated to  
QIAfilter Midi column (Qiagen, Chatsworth, CA). The manufac-  
turer's protocol for washing and eluting the plasmid DNA was  
followed.

50

55

5

10

b. SuperCos: SuperCos plasmid DNA was purchased from Stratagene (La Jolla, CA).

10

3. Preparation of ca. 38-47 kb *Sau3A1* fragments of *S. cellul-  
sum* chromosomal DNA

15

a. *S. cellulorum* chromosomal DNA prepared as described in section A.1.a was partially cleaved with restriction endonuclease *Sau3A1* in a 1000  $\mu$ l reaction volume consisting of 50  $\mu$ g chromosomal DNA, 5 units enzyme (Promega, Madison, WI), 0.006 M Tris-HCl, 0.006 M MgCl<sub>2</sub>, 0.10 M NaCl, and 0.001 M dithiothreitol (pH 7.5) for 5 min at 37°C. The reaction mixture was extracted once with an equal volume of 1:1 phenol:chloroform. After centrifugation, the upper aqueous phase was saved, to which 0.1 vol. of 3 M sodium acetate and 0.6 vol. isopropanol was added. DNA was pelleted by centrifugation for 5 min at 16,000 x g in a microfuge and washed once with 0.5 ml 70% EtOH. After drying in a SpeedVac (Savant Instruments, Farmingdale, NY) for 5 min, the pellet was resuspended in 0.1 ml TE buffer. The DNA was layered on top of a 12 ml 10-40% sucrose gradient prepared in TE buffer and centrifuged at 113,600 x g for 16 hr, 10°C using a Beckman SW40Ti rotor (Beckman Instruments, Palo Alto, CA). Five hundred  $\mu$ l aliquots of the gradient were removed using a pipetor beginning at the top of the tube. Samples (5  $\mu$ l) of the fractions were analyzed by electrophoresis through a 0.5% agarose gel in TAE buffer (0.04 M Trizma base, 0.02 M acetic acid, and 0.001 M EDTA, pH 8.3) containing 0.5  $\mu$ g/ml ethidium bromide for 6 hr at 100 V. Fractions containing DNA fragments of ca. 40-45 kb were identified by comparison to a high molecular weight DNA standard (Life Technologies,

20

10

25

15

30

35

40

45

50

55

5

10

15

Gaithersburg, MD). Sucrose was diluted from the corresponding 0.5 ml fraction by addition of 0.5 vol. TE. Subsequently, DNA was precipitated by addition of 0.1 vol. 3 M sodium acetate and 0.6 vol. isopropanol. DNA was pelleted by centrifugation at 16,000 x g for 10 min in a microfuge. DNA was washed with 0.5 ml 70% EtOH and dried in a SpeedVac with moderate heat for 10 min. Finally, the DNA was resuspended in distilled H<sub>2</sub>O at a concentration of 0.5 mg/ml.

20

25

30

35

40

45

10 b. Alternatively, 10 µg of *S. cellulosum* chromosomal DNA prepared as described in A.1.b was treated with 0.3 U Sau3A1 (New England Biolabs, Beverly, MA) for 1 h at 37°C in 400 µl of the supplier's recommended reaction buffer. Formation of DNA fragments of about 40 kb in size was checked by comparison of the motility behavior with high molecular weight DNA standards after a 0.3% agarose gel electrophoresis. An equal volume of phenol:chloroform (1:1) was added, mixed and centrifuged. The upper aqueous phase was recovered and 0.1 vol. of 3 M sodium acetate and 0.6 vol. of isopropanol were added. After centrifugation, the precipitated DNA was washed twice with 0.5 ml 70% ice cold ethanol and finally air-dried. The DNA fragments were resuspended in 100 µl shrimp alkaline phosphatase reaction buffer and dephosphorylated for 150 min. at 37°C using 2 U shrimp alkaline phosphatase (Amersham Life Science, Cleveland, OH). A phenol:chloroform extraction followed as described above. Finally, the DNA was precipitated by addition of 0.1 vol. 3 M sodium acetate and 0.6 vol. isopropanol, dried, and dissolved in TE buffer.

30

50

55

5

12

#### 4. Preparation of cosmid libraries

10

5

15

20

25

30

35

40

45

50

55

a. Using pFD666: Vector pFD666 was cleaved with restriction endonuclease BamHI in a 0.02 ml reaction volume consisting of 2  $\mu$ g plasmid DNA, 10 units of BamHI (Promega), 0.006 M Tris-HCl, 0.006 M MgCl<sub>2</sub>, 0.05 M NaCl, and 0.001 M dithiothreitol (pH 7.5) for 90 min at 37°C. Five  $\mu$ l of 10x alkaline phosphatase buffer (0.5 M Tris-HCl [pH 9.3], 0.01 M MgCl<sub>2</sub>, 0.001 M ZnCl<sub>2</sub>, 0.01 M spermidine) was added to the reaction followed by alkaline phosphatase (0.01 units/pmol ends; Promega) and distilled H<sub>2</sub>O to a final volume of 0.05 ml. The sample was incubated for 30 min at 37°C and a second aliquot of phosphatase was added. After a further 30 min at 37°C, 0.3 ml of stop buffer (0.01 M Tris-HCl [pH 7.5], 0.001 M EDTA, 0.2 M NaCl, 0.5% SDS) and 0.35 ml of 1:1 phenol; CHCl<sub>3</sub> was added to the reaction. The sample was mixed gently several times by inversion and centrifuged at 16,000 x g for 3 min to separate the phases. The aqueous layer was removed to a new microfuge tube. 0.1 vol. 3 M sodium acetate and 2 vol. 100% EtOH were added and the precipitated DNA pelleted by centrifugation at 16,000 x g for 10 min. Liquid was removed by aspiration and the pellet washed once with 0.5 ml 70% EtOH. The DNA was dried in a SpeedVac and resuspended in TE buffer to 0.5 mg/ml.

Digested, phosphatase-treated pFD666 was ligated to the partially-cleaved chromosomal DNA (see sections A.3.a and B.1.a) in a 0.005 ml reaction consisting of 1  $\mu$ g pFD666, 1  $\mu$ g *S. cellulosum* DNA, 0.03 M Tris-HCl (pH 7.8), 0.01 M MgCl<sub>2</sub>, 0.01 M dithiothreitol, and 0.0005 M adenosine-5'-triphosphate and 1.5 Weiss units of T4 DNA ligase (Promega). The reaction was carried out at room temperature for 2 hr. The entire reaction

5

13

10

15

20

25

30

35

40

45

50

55

mix was packaged into bacteriophage  $\lambda$  in vitro using Packagene extracts (Promega) according to the manufacturer's directions. The entire packaging reaction (0.5 ml) was diluted with 4.5 ml SM buffer (per liter: 5.8 g NaCl, 2 g  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ , 1 M Tris-HCl [pH 7.5], 5 ml 2% gelatin solution). Transfection was performed by adding 10 ml of an overnight culture of *E. coli* DH5 $\alpha$  that had been grown in LB medium with 0.01 M  $\text{MgSO}_4$  and 0.2% maltose to the diluted phage and incubating at 37°C for 20 min. 0.8 ml of LB was added and the cells shaken at 225 rpm for 1 hr at 37°C. Cells were pelleted, resuspended in LB, and spread onto a 150 mm LB + kanamycin agar plate. After 3 d. at 30°C, the colonies were harvested by picking ca. 800 colonies into 2.0 ml LB + kanamycin medium containing 20% glycerol, freezing on dry ice, and storing at -70°C. In addition, six kanamycin-resistant colonies were inoculated into 2 ml LB + kanamycin liquid medium and incubated at 37°C, 250 rpm, for 18-24 hr. Cosmid DNA was prepared using a standard alkaline lysis procedure starting with 1.5 ml of the culture. DNA was digested with restriction endonuclease PstI and samples electrophoresed on a 0.8% TAE agarose gel for 1.5 hr at 100 V. A unique restriction pattern was noted in each sample and the total size of the insert was calculated to be between 40 and 45 kilobases.

b. Using SuperCos: 30  $\mu\text{g}$  of vector SuperCos was digested with XbaI (New England Biolabs, Beverly, MA) for 210 min at 37°C in 100  $\mu\text{l}$  of the recommended reaction buffer. Ten  $\mu\text{l}$  sodium acetate and 60  $\mu\text{l}$  isopropanol was added before the solution was centrifuged for 30 min at 16,000  $\times$  g. The precipitated DNA was washed twice with 500  $\mu\text{l}$  ice cold 70% ethanol. The vector DNA was precipitated and air-dried, dissolved in 135  $\mu\text{l}$  shrimp al-

5

10

15

20

25

30

35

40

45

50

55

kaline phosphatase reaction buffer and treated with 2.5 U shrimp alkaline phosphatase for 150 min. After heat inactivation of the enzyme at 75°C for 20 min, a phenol:chloroform extraction was performed as described in section 1. c. The DNA, resuspended in 100  $\mu$ l BamHI restriction buffer was hydrolyzed with 15 U BamHI (New England Biolabs, Beverly, MA) for 180 min. A phenol:chloroform extraction followed (see section A.3). The SuperCos DNA was precipitated by addition of 0.1 vol 3 M sodium acetate and 0.6 vol isopropanol, centrifuged at 16,000 x g, and resuspended in 50  $\mu$ l TE buffer.

Four  $\mu$ g of digested vector DNA was ligated with 10  $\mu$ g partially hydrolyzed genomic DNA from *S. cellulosum* (as described in section A.3.b) in a final volume of 20  $\mu$ l using 2 U T4 DNA ligase and the appropriate reaction buffer (Gibco BRL, Eggenstein, Germany). The reaction was carried out at 16°C overnight. The reaction mixture was packaged into phage particles using the Gigapack III XL packaging extract kit (Stratagene) according to the manufacturer's protocol. Treatment of packaging reaction mixture and transfection of *E. coli* SURE (Stratagene) was performed as described in 4.a. Transfected cells were concentrated by centrifugation, resuspended in fresh LB medium and distributed on LB agar plates containing 50  $\mu$ g/ml kanamycin. The plates were incubated overnight at 30°C. 1600 recombinant clones were transferred into 96 well microtiter plates filled with 80  $\mu$ l LB medium containing 50  $\mu$ g/ml kanamycin per well and propagated overnight at 30°C. The following day the microtiter plates were used to inoculate a second set of microtiter plates in order to obtain a duplicate of the recombinant clones. Each well of the original set of microtiter plates was supplemented with 80  $\mu$ l 50 % glycerol and the entire plate stored at -70°C.

20 randomly chosen transformants were inoculated into 3 ml LB medium with 50  $\mu\text{g}/\text{ml}^{-1}$  kanamycin and incubated over night at 37°C in order to isolate plasmid DNA using the Qiagen plasmid extraction kit (Qiagen, Hilden, Germany). Restriction fragment analysis of the recombinant cosmids using the restriction endonucleases PstI and BglII indicated that the cosmids contained inserts of approximately 35 to 42 kb in size.

#### B. Construction of a *S. cellulosum* plasmid library

##### 1. Preparation of 8-12 kb fragments of *S. cellulosum* chromosomal DNA.

*S. cellulosum* chromosomal DNA prepared as described in section A.1.a was partially cleaved with restriction endonuclease Sau3AI in a 100  $\mu\text{L}$  reaction volume consisting of 5  $\mu\text{g}$  chromosomal DNA, 5 units enzyme (Promega, Madison, WI), 0.006 M Tris-HCl, 0.006 M  $\text{MgCl}_2$ , 0.10 M NaCl, and 0.001 M dithiothreitol (pH 7.5) for 4 min at 37°C. The digested DNA was electrophoresed through a 11 x 14 cm 0.8% TAE-agarose gel for 18 hr at 17 V. Fragments of 8-12 kb were cut from the gel and purified using the QIAquick Gel Extraction Kit using the manufacturer's protocol (Qiagen).

##### 2. Preparation of the plasmid library

Plasmid pZero2.1 (Invitrogen, Carlsbad, CA) was cleaved with restriction endonuclease BamHI in a 0.02 ml reaction volume consisting of 1  $\mu\text{g}$  plasmid DNA, 10 units of BamHI (Promega), 0.336 M Tris-HCl, 0.006 M  $\text{MgCl}_2$ , 0.05 M NaCl, and 0.001 M di-

5

16

10

15

20

25

30

35

40

45

50

thiothreitol ( pH 7.5) for 20 min at 37°C. 0.08 ml of dH<sub>2</sub>O and 0.1 ml of 1:1 phenol:CHCl<sub>3</sub> was added. The sample was briefly vortexed and centrifuged at 16,000 x g for 2 min. The aqueous layer was removed to a new microfuge tube. 0.1 vol. 3 M sodium acetate and 2 vol. 100% EtOH were added and the precipitated DNA pelleted by centrifugation at 16,000 x g for 10 min. Liquid was removed by aspiration and the pellet washed once with 0.5 ml 70% EtOH. The DNA was dried in a SpeedVac and resuspended in TE buffer to 0.004 µg/ml. Digested pZero2.1 was ligated to the partially-cleaved chromosomal DNA in a 0.01 ml reaction consisting of 0.004 µg pZero2.1, 0.05 µg *S. cellulosum* DNA, 0.03 M Tris-HCl (pH 7.8), 0.01 M MgCl<sub>2</sub>, 0.01 M dithiothreitol, and 0.0005 M adenosine-5'-triphosphate and 1.5 Weiss units of T4 DNA ligase (Promega). The reaction was carried out at room temperature for 2 hr. 0.015 ml dH<sub>2</sub>O and 0.25 ml of 1-butanol were added, the sample vortexed briefly, and centrifuged at 16,000 x g for 10 min. Liquid was aspirated away from the pellet and the sample dried in a SpeedVac for 5 min. The ligated DNA was resuspended in 0.005 ml dH<sub>2</sub>O and mixed with 0.04 ml of electrocompetent *Escherichia coli* DH10B cells (GIBCO/BRL, Gaithersburg, MD). The sample was placed into a pre-chilled 0.2 mm-gap electroporation cuvette and transformed into the bacteria by electroporation using a BioRad Gene Pulser II unit (BioRad, Hercules, CA) at 25 µF and 200 Ω. 0.96 ml SOC medium (0.5% yeast extract, 2% tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl<sub>2</sub>, 20 mM MgSO<sub>4</sub>, 20 mM glucose) was mixed with the cells and transferred to a 1.5 ml microfuge tube. The sample was incubated at 37°C, 225 rpm, for 1 hr. Aliquots of the cells were spread onto an LB agar + kanamycin and incubated at 37°C for 20 hr to estimate the number of transformants obtained. Six kana-

55



5

17

10

mycin resistant colonies were confirmed to contain an insert of the expected size as described in section A.4.a.

C. Identification of cosmids possessing polyketide synthase genes

15

1. Colony blot hybridizations using cosmid library in pFD666:

20

25

30

35

40

45

50

55

A 20 x 20 cm sheet of Duralon UV membrane (Stratagene) was placed on top of a 24.5 x 24.5 cm square bioassay dish containing 250 ml LB agar - kanamycin. An aliquot of the frozen cosmid library in 1 ml LB medium was spread on the filter. The plate was incubated at 37°C for 24 hr. Colonies were replicated onto two fresh filters which were placed onto LB + kanamycin agar medium and incubated at 28°C for 18 hr. Lysis of cells and neutralization of released DNA was performed according to directions that were provided with the filters. The DNA was crosslinked to the filters using a UV Stratalinker 2400 unit (Stratagene) in the auto crosslink mode. Cell debris was removed by placing the filters in a container with a solution of 3 X SSC (20 X SSC contains, per liter, 173.5 g NaCl, 88.2 g sodium citrate, pH adjusted to 7.0 with 10 N NaOH), 0.1% SDS and rubbing the lysed colonies with a Kimwipe. The filters were then incubated at least 3 hr with the same wash solution for at least 3 hr at 65°C. The plasmid library was treated similarly except cells were spread onto a 137 mm circular Duralon UV membrane placed on top of a 150 mm petri dish containing 80 ml LB agar + kanamycin.

For hybridizations, a probe consisting of a 650-base pair (bp) polymerase chain (PCR) fragment representing a portion of a *S. cellulosum* polyketide synthase gene was used. The fragment

was amplified using primers to consensus regions of Type I (macrolide) polyketide synthase (PKS) genes (Swan et al., Mol. Gen. Genetics 242, 358-362 [1994]). A series of sense and anti-sense oligonucleotides were prepared for PCR studies as indicated in the following table 2:

Table 2

Oligo-nucleotide	I. DNA sequence (5' → 3')	Corresponding amino acid sequence
120 (sense)	CGGT (C/G) AAGTC (C/G) AACATCGG	KSNIGHT
121 (anti-sense)	GC (A/G) ATCTC (A/G) CCCTGCGA (A/G) TG	HSQGEIA
122 (sense)	GT (C/G) GACAC (C/G) GC (C/G) TGCTC (C/G)	VDTACSS
123 (sense)	GG (C/G) AC (C/G) AACGC (C/G) CACGT (C/G) A T	GTNAHVI
124 (anti-sense)	CCCTG (C/G) CC (C/G) GGGAA (C/G) ACGAA	FVFPGQG

10

The selection of C or G where necessary in the third position of a codon reflects the very high overall G + C content of *S. cellulosum* (ca. 70%). Conditions for PCR were as follows: 0.01 M Tris-HCl (pH 9.0), 0.05 M KCl, 0.003 M MgCl<sub>2</sub>, 0.1% Triton X-100, 200 μM of each primer, 2.5 U Taq DNA polymerase (Promega), 5.0% dimethyl sulfoxide (Sigma), and 0.01 μg of *S. cellulosum* chromosomal DNA in a 0.25 ml reaction volume. Reactions were

5

19

10

15

20

25

30

35

40

45

50

55

carried out in a Perkin-Elmer Model 480 Thermocycler (Perkin-Elmer Corporation, Foster City, CA) under the following conditions: 94°C, 1 min; 50°C, 1 min, 72°C, 1.5 min for a total of 30 cycles. Each possible combination of sense and anti-sense primers were tried. A 650-bp and 350-bp fragment was amplified using cligos 120 + 124 and 123 +124, respectively. The sequence of the fragments were determined using the ALFexpress AutoRead kit to fluorescently label the DNA, which was analyzed on an ALFexpress sequencing apparatus (Pharmacia). The data indicated both PCR fragments possessed significant homology to polyketide synthase genes of Type I antibiotics. The 650-bp fragment was chosen for hybridization experiments.

The fragment was labeled with <sup>32</sup>P-dCTP using the NEBlot kit (New England Biolabs, Beverly, MA) and purified on a Bio-Spin 6 column (BioRad, Hercules, CA.). Duplicate blots were pre-hybridized in 3 X SSC (1 X SSC contains 0.15 M sodium chloride and 0.015 M sodium citrate, pH 7.0), 4 X Denhardt's solution (100 X is 2% Ficoll [Type 400], 2% polyvinylpyrrolidone, and 2% bovine serum albumin [Fraction V]), and 100 µg/ml sheared, denatured salmon sperm DNA; all reagents purchased from Sigma Chemicals, St. Louis. The labeled DNA was heated in a boiling water bath for 5 min to denature the strands, cooled on ice, and added to the pre-hybridization solution. The filters were incubated for at least 18 hr in a roller bottle hybridization oven. They were transferred to new bottle, then washed two times in 2 X SSC, 0.1% SDS at 70°C for 30 min (moderate stringency). The membranes were placed on Whatman 3MM paper to remove excess liquid, covered with Saran Wrap, and exposed to autoradiography film (Kodak X-OMAT LS) with two intensifying

5

20

screens. The cassette was placed at -70°C and developed at appropriate intervals.

10

Approximately 100 colonies were seen to have hybridized on the duplicate filters. Fourteen of these were isolated from the master plate and grown in 4 ml LB + kanamycin medium for 20-24 hr, 37°C, 250 rpm. Plasmid DNA was prepared using the standard alkaline lysis method and digested with restriction endonuclease PstI. The digested DNA was electrophoresed on a 0.8% agarose gel in TAE for 3 hr at 100 V. Fragments were transferred to Duralon UV using the VacuGene XL vacuum blotting unit (Pharmacia) and the recommended alkaline denaturation protocol. Hybridization with radioactively-labeled PCR fragment and washing were carried out as described above. Two prominent types of cosmids were observed; one contained PstI fragments of ca. 7.0, 5.0, and 1.1 kb (pEPOcos6 and pEPOcos7) that hybridized to the probe; the other type had fragments of ca. 6.0 and 3.6 kb (pEPOcos8 and pEPOcos13) which were homologous to the probe. Restriction analysis confirmed that cosmids showing identical hybridization patterns had identical or overlapping inserts. PCR reactions using primers representing consensus sequences of Type I PKS genes were performed using the isolated cosmid DNA as template under conditions described above, except ca. 0.01 µg of cosmid DNA was included as template. Cosmids pEPOcos6 and pEPOcos8 amplified the 650-bp fragment seen when oligonucleotides 120 + 124 were used, while pEPOcos8 and pEPOcos13 supported amplification of an 1100-bp PCR fragment with oligos 122 and 124. The latter fragment was sequenced and confirmed to possess strong similarity to Type I PKS genes. These data confirm that the recombinant cosmids are related to each other and that all contain PKS-like genes.

50

55

5

21

10

## 2. Colony blot hybridizations of plasmid library in pZero2.1:

15

20

25

A 137-mm circle of Duralon UV membrane was placed on top of a 150-mm containing 75 ml LB agar + kanamycin. An aliquot of the plasmid library (representing ca. 2,000 recombinant colonies) in 0.5 ml LB medium was spread on the filter. The plate was incubated at 37°C for 20 hr. Colonies were replicated onto two fresh filters which were placed onto LB + kanamycin agar medium and incubated at 37°C for 6 hr. The filters were processed for hybridization as described in Section C.1. Out of 8 positive colonies detected, one contained a plasmid with a DNA region not encoded by either pEPOcos6 or pEPOcos8. This plasmid, called Sau4, was characterized in more detail.

## 3. Colony blot hybridizations of cosmid library in SuperCos:

30

35

40

45

50

The recombinant *E. coli* clones from the microtiter plates (see section 4. b) were used to produce two identical sets of hybridization filters in order to identify cosmids carrying PKS and PS genes. The recombinant clones were spotted onto 2 sets of 22 x 22 cm LB agar plates containing 50 µg/ml kanamycin. Each plate contained 384 clones therefore representing 4 microtiter plates. The clones were incubated at 30°C overnight. After pre-cooling for approximately 3 h at 4°C, 20 x 20 cm Hybond N<sup>+</sup> Nylon membranes (Amersham, Braunschweig, Germany) were placed onto the agar surfaces. After 2 min. the membranes were removed and placed for 15 min. on Whatman 3 MM paper (Whatman paper Ltd., Maidstone, England) soaked with denaturation solution (0.5 N NaOH, 1.5 M NaCl) before they were transferred onto Whatman 3 MM paper saturated with neutralization solution (1 M Tris-HCl, pH 7.5, 1.5 M NaCl). Subsequently the membranes were

55

5

22

10

15

placed onto Whatman 3 MM paper soaked with 2 X SSC (0.3 M NaCl, 0.03 M sodium citrate, pH 7.2) for 10 min. The membranes were baked for 40 min at 85°C. Then, each membrane was overlaid with 5 ml Proteinase K solution (2 mg/ml Proteinase K in 2 x SSC) and incubated at 37 °C for 90 min. Finally, cell debris was removed by wiping the membranes with a Kimwipe pre-wetted with 2 X SSC.

20

25

30

As we were seeking in particular to identify biosynthetic pathways containing both PKS and PS genes, the following hybridization strategy was taken: The screening was initially focused on ketosynthase domains from type I PKSs and on the adenylation domain from PSs. Target-specific primers were used to amplify DNA fragments of the corresponding genes from chromosomal DNA of *S. cellulosum* by PCR. The fragments obtained were then cloned, sequenced and the deduced amino acid sequence compared to known ketosynthase and adenylation domains of PKS and PS, respectively. In a second step these PCR fragments were used as gene probes to detect recombinant cosmids of the *S. cellulosum* cosmid library.

35

40

Oligonucleotides based on conserved amino acid sequences of ketosynthase domains from various type I PKS were optimized for myxobacterial DNA by comparison to a known myxobacterial biosynthetic gene cluster (Schupp et al., J. Bacteriol. 177, 3673-3679 [1995]) resulting in primer

KS1Up (5'-

C/A)GIGA(A/G)GCI(A/C/T)(A/T)I(C/G)(C/A)IATGGA(C/T)CCICA(A/G)CAI(A/C)G-3') and

45

KSE1 (5'-GG(A/G)TCICCA(A/G)I(G/C)(T/A)IGTICIGTICC(A/G)TG-3').

PCR-primers TGD (5'-

T(A/T)(C/T)CGIACIGGIGA(C/T)(C/T)(G/T)IG(G/T)ICG-3') and

50

55

5

23

LGG (5'-

10

A(A/T)IGA(A/G)(G/T)(G/C)ICCI(A/G)(A/G)(G/C)I(A/C)(A/G)AA(A/G)  
)AA-3')

15

20

25

30

35

40

45

50

55

directed to genes encoding adenylation modules have been described by Turgay et al. (Pept. Res. 7, 238-241 [1994]). PCR reaction mixtures with a final volume of 25  $\mu$ l contained 0.1  $\mu$ g template DNA, 0.2 U Taq DNA-polymerase (Gibco BRL, Eggenstein, Germany), 5  $\mu$ mol dNTP, 5% dimethyl sulfoxide (Sigma), 1.5 mM  $MgCl_2$ , 25 pmol of each primer and the appropriate reaction buffer supplied by Gibco BRL. Chromosomal DNA of *S. celluloseum* was used as template. Additionally, chromosomal DNA of *Myxococcus fulvus* was used with PS primers. Reactions were carried out in an Eppendorf Mastercycler Gradient (Eppendorf, Germany) using the following conditions: denaturation 30 s at 97°C, annealing 30 s at 55°C, extension 60 s at 72°C for a total of 30 cycles. The formation of ca. 700 bp fragments using the KS primers and of ca. 350 bp fragments with the PS primers were confirmed by 0.8% agarose gel electrophoresis. Fragments of independent PCR reactions were ligated into vector pCR2.1TOPO using the TOPO TA Cloning kit (Invitrogen, Leek, The Netherlands) according to the manufacturer's protocol and transformed into *E. coli* XL1-Blue. Sequencing of the resulting plasmids and analysis of the deduced amino acid sequence revealed three different KS fragments, designated pM008.4, pM008.6, pM008.7, one PS fragment (pAPs1) corresponding to *S. celluloseum* and one PS fragment (pDPs1) obtained with chromosomal DNA of *M. fulvus*. The PCR fragments were re-isolated by digestion with *Eco*RI from the plasmids pM008.4, pM008.6, and pM008.7, labeled, pooled and used as gene probes in hybridization experiments as described

5

24

below. The same procedure was performed with the PS fragments of pAPs1 and pDPs1.

10

15

20

25

30

Hybridization with PKS and PS specific DNA probes (see above) was carried out using the DIG nonradioactive labeling and detection kit (Boehringer Mannheim, Germany) and performed according to the supplier's manual using buffer containing 50% formamide. The membranes were hybridized in plastic bags containing approx. 10 ml of hybridization solution at 39°C overnight. Unspecific binding of probes was removed by 2 wash steps with 2 x SSC, 0.1% SDS at room temperature for 20 min. and one stringent wash step with 0.5 x SSC, 0.1% SDS at 60°C for 20 min. Detection of hybridizing DNA fragments was performed with the above mentioned system according to the manufacturer's protocol using CSPD as chemiluminescent substrate. The signals were recorded by exposure of the treated membrane to Hyperfilm ECL (Amersham Life Science, Little Chalfont, England) which was developed in appropriate time intervals.

35

40

45

71 signals were detected with the PKS specific gene probe. On the duplicate filters 35 signals were obtained with the PS specific gene probe of which 7 were already known from the PKS hybridization experiment. These recombinant cosmids harbored PKS- and PS-encoding genes. In order to corroborate these results PCR experiments were performed with DNA of the 7 recombinant cosmids as template and PKS (KS1Up, KSD1) and PS specific primers (TGD, LGG) generating fragments of the expected size of approx. 700 bp and 350 bp, respectively (primers and reaction conditions see above).

50

A comparison of the restriction fragment patterns of the DNA from the 7 recombinant cosmids carrying PKS and PS genes digested by BamHI facilitated an arrangement of the cosmids in

55



5

25

10

3 groups. They were represented by cosmids designated A2 and A5. The remaining group was represented by pEPOcos6. Therefore, A2 and A5 represented good candidates for further DNA sequence analysis because they carry both PKS and PS genes.

5

15

#### D. Random "shotgun" sequencing of recombinant cosmids and plasmids

20

##### 1. Library construction

10

25

30

a. pEPOcos6, pEPOcos8, A5, and Sau4: pEPOcos6 and pEPOcos7 were sequenced to completion, and contiguous sequence data and analysis for these overlapping cosmids is presented below for the "cos6 region" (cf. claims 7 and 9). Sequencing of cosmid A5, pEPOcos8 and plasmid Sau4 was taken to the point of large contiguous sequences (contigs) representing the *S. cellulosum* insert; sequence and analysis presented below (cf. claims 10 to 15).

35

40

45

Randomly sheared libraries were constructed for cosmids and plasmids of interest using a protocol similar to that of of Fleischmann et al., 1993 (Science 269, 496) and modified in Fraser et al., 1995 (Science 370, 397). Briefly, Qiagen-column purified cosmid DNA (~10 µg) was sheared to a size of approximately 2 kb and the DNA end-repaired using BAL31 nuclease. The DNA was gel-purified after electrophoresis through a 0.75% low-melting temperature agarose gel containing 0.5 µg/ml ethidium bromide in 1X TAE buffer run at 80 V for 2 hours. The volume of the low-melt agarose gel slice was estimated by adding the gel slice to a microfuge tube and weighing, then 0.1 vol. of 3 M sodium acetate (pH 7) was added and the agarose incubated at 60°C. The temperature was equilibrated to 37°C, and DNA ex-

50

55

5

26

10

15

20

25

15

30

35

40

25

tracted twice using an equal volume of buffered phenol (Life Technologies). The aqueous phase was transferred and extracted once with an equal volume of chloroform, then ethanol precipitated by the addition of 2 vol. cold 100% ethanol. DNA was concentrated by spinning at 16,000 x g in a microcentrifuge. The DNA pellet was washed with 1 ml 70% ethanol and resuspended in 100  $\mu$ l of 0.1X TE. The DNA was ligated to SmaI-digested, phosphatase-treated pUC18 vector (Pharmacia), and single insert recombinants isolated by gel-purification of the band containing vector plus a single insert, followed by T4 polymerase polishing, and a final intramolecular ligation of the vector-plus-single-insert DNA. This final ligation represents a library of highly random ca. 2 kb fragments that was used for shotgun sequencing of the ca. 40 kb cosmids or ca. 10 kb plasmids.

b. Cosmid A2: Cosmid DNA with inserts of *S. cellulorum* was isolated by an alkaline lysis procedure and purified with Macherey Nagel columns (Macherey und Nagel GmbH und CoKG, Düren, Germany) using manufacturer's recommendation. Purified Cosmid DNA was sonicated, end-repaired using T4 DNA Polymerase (Boehringer Mannheim, Germany). After gel-purification fragments of a size of approximately 2 kb were ligated into SmaI-digested, phosphatase-treated pTZ18R vector (Pharmacia). The ligation represents a library of highly random ca. 2 kb fragments that was used for shotgun sequencing of the ca. 40 kb cosmid.

## 2. Sequencing and assembly

45

a. pEPOcos6, pEPOcos8, Sau4, and A5: DNA (1  $\mu$ l of 100  $\mu$ l total in the library) was transformed into *E. coli* by electroporation (20  $\mu$ l of Electromax DH10B cells from Life Technolo-

50

55

5

27

10

15

20

25

gies) and cells spread onto LB plates containing 50 µg/ml ampicillin. After growth overnight at 37 °C, transformants (ca. 300-3000 CFU total) were transferred to 96-well growth blocks and shaken overnight at 37°C in 1.3 ml LB medium with 50 µg/ml ampicillin. Templates were prepared from these cells by an alkaline lysis procedure (Qiagen QiaQuick Turbo Prep) to yield purified, double-stranded plasmid DNA. Cycle-sequencing of the plasmid templates was performed using universal forward and reverse primers and BigDye Terminator sequencing kits (Applied Biosystems), using the manufacturer's recommendations, then resolved using an ABI377 automated sequencer. Sequences were edited using Phred, then assembled into larger contiguous sequences using Phrap (Phil Green, University of Washington, St. Louis, MO).

15

30

35

40

45

b. Cosmid A2: DNA (1 µl of 20 µl total in the ligation) was transformed into *E. coli* DH10B by electroporation and cells were spread onto LB agar medium containing 50 mg/ml ampicillin. After growth for 18 hr at 37°C, transformants were transferred to 96-well growth blocks and shaken overnight at 37°C in 1.3 ml 2x YT medium with 50 mg/ml ampicillin. Templates were prepared from these cells by an alkaline lysis procedure (Qiagen QiaQuick Turbo Prep) to yield purified, double-stranded plasmid DNA. Cycle-sequencing of the plasmid templates was performed using universal forward and reverse primers and Big Dye Terminator sequencing kits (PEBiosystems) or Thermo Sequenase fluorescent labelled primer cycle sequencing kit (Amersham Pharmacia Biotech) using the manufacturer's protocols. In the shotgun phase of a cosmid, identical amounts of samples were sequenced either by dye-primer or dye-terminator chemistries (Pharmacia, PE Biosystems). Data were collected using Licor and ABI 377

50

55

5

28

10

automated sequencers and assembled with the GAP4 program (Bonfield, Smith, Staden, Nucl. Acids Res. 23, 4992-4999 [1995]). Gaps were closed using custom made primers (MWG-Biotech) on plasmid templates or PCR products in combination with dye-terminators.

15

## E. Bioinformatic Methods

20

### 1. Open reading frame (ORF) identification

10

ORFs were identified in the pEPOcos6 region using the OMIGA 1.1.2 (GCG 0.4D) program from Oxford Molecular Limited. Default values were used (Standard genetic code, all ORFs over 50 bases) to generate ORFs; analysis of these results lead to the list of 14 highest quality ORFs as defined in claim 9. Other ORFs, genes, or genetic elements may be found in the pEPOcos6 insert that have not yet been annotated. In addition to hand-editing of the OMIGA-generated data, the MAGPIE automated genome analysis tool:

25

30

(<http://genomes.rockefeller.edu/magpie/magpie.html>)

35

20

was used to identify genes for all the sequenced cosmids and plasmids. ORFs identified in this manner are presented as both nucleotide and peptide files below.

40

For cosmids A2 and A5, ORFs have been identified within the DNA sequences of A5 (contigs 10, 11, 12) and of A2 using the FramePlot analysis program from Ishikawa and Hotta (FEMS Microbiol. Lett., 174, 251-253 [1999] public available under <http://www.nih.go.jp/~jun/cgi-bin/frameplot.pl>) which is based on positional base preference in codons typical for organisms having genomes with a high G + C content (Bibb et al., Gene 30,

45

50

55

157-166 [1984])). Default parameters using ATG and GTG as start codons were used. The deduced amino acid sequence of predicted ORFs were compared with protein databases (GenBank, CDS translations, PDB, SwissProt, PIR, PRF) using BLASTP (Altschul et al., Nucleic Acids Res., 25, 3389-3402 [1997])). Additionally, high scoring amino acid sequences were analyzed using the Pfam program [<http://www.sanger.ac.uk/Software/Pfam/>], which identified specific domain structures of the submitted proteins (Bateman et al. Nucleic Acids Res., 27, 260-262 [1999])).

## 2. BLAST searches

BLASTP2 similarity searches were performed using the peptide files from the above ORF identification strategy as query sequences. Searches were performed using the in-house Bioinformatics BLASTP2 (Version: BLASTP 2.0a19MP-WashU) web page at the Bristol-Myers Squibb Pharmaceutical Research Institute (allows BlastN2, BlastP2, BlastX2, TblastN, and TblastX searches). In addition, peptide files generated by the MAGPIE analysis were automatically searched using a FASTA algorithm.

## 3. Best match and probable identification

Analysis of the BLASTP2 and FASTA output led to an assignment of a best match and probable function. The best match was usually the top scoring match, although sometimes another match was given because it was a more relevant homolog, or no match was found with a significance greater than  $>e-4$ . Probable function represents the best estimate of function given the initial analysis of the BLAST data and the published literature regarding the best match, and may not necessarily represent the true function of the gene product (hypothetical proteins are of un-

5

30

10

15

known function). A higher probability score indicates a higher likelihood that the probable function corresponds to that of the best match; e.g., the polyketide synthase matches are all above e-100, and given the very high significance scores are presumed to function as polyketide synthases (as are the high scoring peptide synthetases).

The following is a summary of the sequence data from the pEPOcos6 region, pEPOcos8, A5, Sau4 and A2.

20

10 a. Data from pEPOcos6 region:

25

Summary: A large PKS/PS cluster spanning multiple cosmids. An IS element (designated IS-Sc1 here) is found in the cluster - this may be a potential tool for genetic analysis of *Sorangium*.

15

30

Statistics: Sequence was assembled from over 2000 random sequences (forward and reverse reads of the ca. 2 kb cloned fragments derived).

20

35

47,713 nucleotides of contiguous sequence (no pFD666 vector included)

40

25

DNA sequence data are as defined in claim 7.

45

Note: pEPOcos6\_ORF7 sequences (cf. claim 9): the predicted N-terminus of ORF7 shows 145 nucleotide overlap with ORF6.

30

Note: pEPOcos6\_ORF8 sequences (cf. claim 9): >pEPOcos5\_ORF8.seq ("ORF9\_up" in Fig.2)

50

55

5

31

10

57.3% G+C

15

Table 3 shows ORF data summary. Note: pEPOcos6\_ORF1.seq is truncated at its 5' end; correspondingly pEPOcos6\_ORF1.pep is truncated at its N-terminus.

**b. Data from pEPOcos8 region:**

20

Summary: Two PKS genes found on a cosmid. A Tn1000 insertion is also found (occurred during *E. coli* propagation). No peptide synthetase genes were found; one P450 hydroxylase was identified.

25

30

Statistics: 1952 random sequence reads from the pEPOcos8 library were assembled using phrap, with 1024 of the sequences assembling into 57 contigs. 12 of these contigs were chosen (totaling 56,537 bp) which each contained >6 reads and consisted of about 1000 bp or more. The sequences of these 12 contigs and the associated ORFs are given below.

35

DNA sequence data from contigs are as defined in claim 10. Table 4 shows more data.

40

pEPOcos8 protein data are as defined in claim 11, i.e. for selected ORFs (polyketide synthase, peptide synthetases, or ORFs with high similarity to known genes).

45

50

55

5

32

c. Data from cosmid A5 insert:

10

Summary: A cluster of PKS and PS genes found on the cosmid. Other genes possibly involved in this secondary metabolite production include a downstream lipxygenase gene highly similar to eukaryotic orthologs.

15

20

Statistics: 880 random sequence reads from the A5 library were assembled using phrap, with 530 of the sequences assembling into 12 contigs. 3 of these contigs were chosen (totaling 41,556 bp) which each contained >100 reads and consisted of about 9000 bp or more. The sequences of these 3 contigs and the associated ORFs are given below.

25

DNA sequence data from contigs are as defined in claim 12. Table 5 shows more data.

30

Protein sequence data from selected A5 ORFs are as defined in claim 13.

20

35

d. Data from plasmid Sau4 insert:

40

Summary: Insert contains PKS genes on two large contigs - most similar to the soraphen PKS gene from Sorangium.

25

45

Statistics: 565 random sequence reads from the Sau4 library were assembled using phrap, with 84 of the sequences assembling into 18 contigs. 2 of these contigs were chosen (totaling 6596 bp) which each contained >10 reads and consisted of

50

55



about 1000 bp or more. The sequences of these 2 contigs and the associated ORFs are given below.

DNA sequence data from plasmid Sau4 contigs are as defined in claim 14. Table 6 shows more data.

Protein sequence data from selected plasmid Sau4 ORFs are as defined in claim 15.

e. Data from cosmid A2

Table 7 shows ORF data summary

F. Construction of suitable recombinant expression vectors

1. Expression in Myxobacteria

Heterologous expression of the ORFs shown in Figure 1 is performed by using a derivative of plasmid pSUP102 (Simon, R., Priefer, U., Pühler, A., Methods in Enzymology (1986), vol. 118, pp. 643-659). In this plasmid the gene for chloramphenicol resistance is changed for a cassette comprising the gene for streptomycin resistance and the promoter element of the Tn5 transposon. Short homologous genomic DNA segments from the host organism are ligated with the DNA sequences of Figure 1 and with efficient regulatory elements into, for example, the EcoRI restriction site of the vector. Following amplification of the vectors in *Escherichia coli* the DNA is transferred by electroporation of the host cells or by conjugation with *Escherichia coli* S17-I (Simon, R., Priefer, U., Pühler, A., Biotechnology (1983), vol. 1, pp. 784-791).

5

34

10

By means of the tetracycline or streptomycin resistance, respectively, mediated by the vector the host cells are checked for integration of recombinant plasmid DNA into the chromosome by homologous recombination.

5

15

## 2. Expression in Streptomyces cells

Heterologous expression of the ORFs shown in Figure 1 is performed by using bifunctional Streptomyces-Escherichia coli cosmids pKU206 and pOJ466.

20

10

## 3. Expression in Escherichia coli cells

Heterologous expression of the ORFs shown in Figure 1 is performed by using "bacterial artificial chromosomes", cosmids (for example Supercos, Stratagene GmbH, Heidelberg) and T7 expression systems (Stratagene GmbH, Heidelberg; New England Biolabs Schwalbach, FRG). Expression of recombinant enzymes occurs in Escherichia coli cells constitutively expressing phosphopantetheinyl transferase required for the formation of holoenzyme polyketide synthetases and polypeptide synthetases.

25

15

30

35

40

45

50

55

Table 3. pEP0cos6 region gene annotation summary (continued).

GENE	5'	3' # BP	#AA	MM (DA)	BEST MATCH [SCORE]	PROBABLE IDENTIFICATION
ORF1	1 <sup>a</sup> 731	732 <sup>b</sup> 244	26218	ACOA_BACSU [e-22]	acyl-CoA dehydrogenase	
ORF2	731 2275	1545 515	56317	AF047828 [e-94]	peptide synthetase	
ORF3	2301 3356	1058 352	39092	U57065 ZnaR [e-29]	antibiotic resistance gene	
ORF4	3412 4050	639 213	24819	(no match)	hypothetical protein	
ORF5	4248 12101	7854 2618	283282	PKSK_BACSU [e-190]	polyketide synthase	
ORF6	12097 14322	2226 742	80545	283857 ppsE [e-111]	polyketide synthase	
ORF7	14178 <sup>c</sup> 16451	2274 758	82007	(no match)	hypothetical protein	
ORF7.1 <sup>d</sup>	15866 14196	1671 557	61320	Y4HP_RHISN [e-28]	hypothetical protein	
ORF7.2 <sup>d</sup>	16507 16154	354 118	13102	Y4HO_RHISN [e-13]	hypothetical protein	
ORF7.3 <sup>d</sup>	16815 16507	309 103	11079	Y4RG_RHISN [e-13]	hypothetical protein	
ORF8	16830 17630	801 123	29633	PKSL_BACSU [e-2]	polyketide synthase	
ORF9	17628 25244	7617 2539	279286	pir1873015 [e-160]	polyketide synthase	
ORF10	25235 27877	2643 881	97101	AF047828 [e-111]	peptide synthetase	
ORF11	27867 32498	4632 1544	165084	AF091251 [e-167]	polyketide synthase	
ORF12	32498 39922	7425 2475	267116	AF040570 [e-223]	polyketide synthase	
ORF13	40031 45559	5529 1843	71258	AF091251 [e-119]	polyketide synthase	
ORF13.1	45599 46018	420 140	14590	(no match)	hypothetical protein	
ORF14	46015 47577	1563 521	55671	PKN1 MYXXA [e-34]	Serine/Threonine-Protein kinase	

<sup>a</sup> The predicted ORF1 gene and gene product is truncated due to cloning of the DNA into the cosmid vector.

<sup>b</sup> 731 is the last nucleotide of the last amino-acid-encoding codon; 732-735

5

36

10

15

20

25

30

35

40

45

50

55

is TGA (termination codon). Termination codons have been excluded in the present annotation.

\* note: the predicted N-terminus of ORF7 shows 145 nucleotide overlap with ORF6

\* note: ORFs on opposite strand - make up a transposable element from bp 16863-14130 (2733 bp, 11 bp terminal inverted repeat) that is similar to IS1131 from *Agrobacterium tumefaciens* (IS-66 like element, 2773 bp, 4 ORFs, 11 bp inverted repeat)

Table 4. p8POcos8 assembly analysis summary (continued).

## a. p8POcos8 assemblies

ASSEMBLY	SIZE (bp)
Contig43	1017
Contig44	1246
Contig48	978
Contig49	1969
Contig50	2877
Contig51	2319
Contig52	1883
Contig53	4871
Contig54	7257
Contig55	5021
Contig56	10945
Contig57	16154

## b. selected ORFs from Contig 56 and 57 of p8POcos8

gene	5'	3'	θ	bp	faa	best match (score)	probable identification
Contig56_003	3	8675	8673	2890	ERY2_SACER (e-300+)	polyketide synthase	
Contig56_027	10784	8682	2103	700	pir1160218 (e-300+)	transposon Tn1000 (E. coli)	
Contig57_001	92	1210	1119	372			
Contig57_002	222	7001	6780	2259	AL021899 pks12 (e-267)	polyketide synthase (M. tuberculosis)	

38

Contig57_026	7424	6975	450	149		
Contig57_027	7116	8375	1260	419	CPKK_SACER [e-50]	Cytochrome P450 monooxygenase
Contig57_030	9074	9337	738	245		
Contig57_031	8412	9080	669	222		
Contig57_032	8432	9667	1236	411		
Contig57_033	9724	8600	1125	374		
Contig57_037	10621	9755	867	288		
Contig57_039	11085	10618	468	155		
Contig57_040	10687	11091	405	134		
Contig57_041	11887	10904	984	327		
Contig57_043	11520	12998	1479	492	AF072709 [e-42]	unknown ORF (S. lividans)
Contig57_044	13730	12018	1713	570		
Contig57_047	13093	13797	705	234		
Contig57_048	14064	13138	927	308		
Contig57_049	14371	13535	837	278		
Contig57_051	13900	14394	495	164		
Contig57_052	14036	14440	405	134		
Contig57_053	14678	15715	1038	345		
Contig57_056	15211	15900	690	229		

Table: 5. AS assembly analysis summary (continued)  
a. pEPOcos8 assemblies

contig	bp	ORF	5'	3'	bp	G+C content pos.3	best match [score]	probable function	identified domains using Pfam
10	9435	1	2861	582	2280	88.6 %	AAC44128 (e-170)	peptide synthetase [Salivarycin]	AMP-binding aa 68-512; pp-binding aa 631-697
		2	3525	2802	624	87.0 %	no match	OmpA like	
		3	6420	4195	2226	85.1 %	P38370 (e-04)	unknown	
		4	7734	7135	600	98.0 %	P39760 (e-21)		
		5	9434	7950	>1495	96.2 %	no match		
11	18023	1	2526	751	1776	97.1 %	CA836518 (e-134)	peptide synthetase	AMP-binding aa 118-520;
		2	3588	2501	1086	94.5 %	CA836516 (e-12)	peptide synthetase	DUF4 (domain of unknown function, found to the carboxy side of pp-binding sites) aa 5-360;
		3	5210	3579	1632	97.8 %	SS3990 (e-102)	peptide synthetase [Pyoverdine]	pp-binding aa 1-30; DUF4 aa 51-442; conflict ca. 20 aa at N-terminus are missing (conserved serine residue) conflict!
		4	6261	3811	2649	64.4 %	CAA11039 (7e-65)	polyketide synthase [Rifamycin]	ketosyl-synthase aa 14-439;
		5	6087	6258	1830	94.8 %	CAB05094 (e-144)	polyketide synthase [Phenolphthalein]	AMP-binding aa 371-758; pp-binding aa 840-904; DUF4 aa 937-1318; AMP-binding aa 1438-1825; pp-binding aa 1907-1071; DUF4 aa 1969-2388; AMP-binding aa 2485-2889; pp-binding aa 2872-3004; pp-binding aa 3027-3046;
		6	17320	6090	9231	92.8 %	AAD04757 (e-180)	peptide synthetase [Mycob.]	DUF4 aa 43-476;
12	15898	1	1	1545	>1545	97.5 %	AAC68816 (4e-74)	peptide synthetase [FK506]	
		2	2883	1549	1335	94.4 %	P43492 (e-07)	cytochrome P450 enzyme	
		3	4659	2911	1749	92.5 %	NP 001130 (5e-56)	lipoygenase	
		4	5896	7066	1191	95.2 %	no match		
		5	7094	7822	729	90.1 %	no match		
		6	8974	7843	1032	87.8 %	no match		
		7	12001	11252	750	92.8 %	no match		
		8	13533	12181	1353	93.3 %	no match		
		9	15897	13981	>1917	93.9 %	CAA19149 (e-08)	regulator	

43368

Table 5.

## b. selected ORFs from cosmid A5

gene	5'	3'	bp	kaa	best match [score]	probable identification
Contig10_001	2861	582	2280	759	U24657 saframycin (e-155)	polyketide synthase
Contig10_006	2880	3323	444	147		
Contig10_007	3876	2902	975	324		
Contig10_008	2953	3957	1005	334		
Contig10_009	4357	3299	1059	352		
Contig10_013	4180	6585	2406	801		
Contig10_016	7033	5600	1434	477		
Contig10_017	7043	6138	906	301		
Contig10_018	6500	7219	720	239		
Contig10_019	6954	7295	342	113		
Contig10_020	6982	7797	816	271		
Contig10_021	7819	7040	780	259		
Contig10_023	7415	8029	615	204		
Contig10_024	7794	9435	1642	548		
Contig10_027	9435	8806	630	209		
Contig11_001	770	417	354	117		
Contig11_002	2526	751	1776	591	AL035640 (e-113)	peptide synthetase
Contig11_004	1033	2787	1755	594		
Contig11_005	3500	1977	1524	507		
Contig11_007	3586	2501	1086	361		
Contig11_008	2507	3814	1308	435		
Contig11_011	5213	3579	1635	544	AF047828 syringomycin (e-86)	peptide synthetase



5

10

15

20

25

30

35

40

45

50

55

41

Contigl1_012	6459	3811	2649	882	AJ223012 (e-52)	rifamycin	polyketide synthase
Contigl1_016	6511	5210	1302	433			
Contigl1_018	5551	6735	1185	394			
Contigl1_021	9899	6258	3642	1213	283857	ppsE M. tuberc	[e-129] polyketide synthase
Contigl1_026	17329	8090	9240	3079	AF047828	syringomycin (e-300+)	peptide synthetase
Contigl1_048	16733	17365	633	210			
Contigl1_049	17397	17723	327	108			
Contigl2_001	1	1545	1545	514	AF082100 (e-63)	(Streptomyces)	peptide synthetase
Contigl2_002	1368	1	1368	456			
Contigl2_003	3	1655	1653	550			
Contigl2_005	2317	1361	957	318			
Contigl2_006	2883	1549	1335	444			
Contigl2_007	1777	4926	3150	1049			
Contigl2_009	4659	2911	1749	582	LOX5_MOUSE (e-54)		ARACHIDONATE 5-LIPOXYGENASE
Contigl2_011	4523	5065	543	180			
Contigl2_012	4638	5231	594	197			
Contigl2_013	4942	5520	579	192			
Contigl2_014	6056	5541	516	171			
Contigl2_015	5765	6373	609	202			
Contigl2_016	5896	7086	1191	396			
Contigl2_017	7095	5899	1197	398			
Contigl2_018	5955	7331	1377	458			
Contigl2_020	7549	7010	540	179			
Contigl2_021	7094	7822	729	242			
Contigl2_022	7995	7318	678	225			
Contigl2_023	8408	7716	693	230			
Contigl2_024	7916	9550	1635	544			

5

42

10

15

20

25

30

35

40

45

50

55

Contig12_029	10751	9732	1020	339
Contig12_030	9795	11132	1338	445
Contig12_033	10880	12124	1245	414
Contig12_035	12184	11252	933	310
Contig12_036	12079	14098	2020	674
Contig12_039	14098	12248	1851	616

Table 6. Sau4 assembly analysis summary.

## a. plasmid Sau4 assemblies

<u>Assembly size (bp)</u>	
Contig17	2581
Contig18	4015

## b. selected ORFs from cosmid A5

gene	5'	3'	# bp	lan	best match (score)	probable identification
Contig17_001	2485	1	2485	829	U24241 Sorangium (e-213)	polyketide synthase
Contig18_002	2	1510	1509	502	U24241 Sorangium (e-105)	polyketide synthase
Contig18_010	1494	4015	2522	841	U24241 Sorangium (e-245)	polyketide synthase

Table: 7. ORF data summary from A2 insert

ORF	5'	3'	bp	G+C content pos.3	best match [score]	probable function	identified domains using Pfam
1	1666	1	>1666	94.4 %	P54744 [5e-37]	regulation [serine/threonine protein kinase]	kinase aa 47-294;
2	1805	3338	1734	90.7 %	no match	acyl-IRNA synthetase	IRNA-synthase aa 27-694;
3	6100	3398	2703	96.0 %	CAA15124 [2e-180]	monooxygenase	oxidoreductase FAD/NAD-binding domain aa 110-227;
4	7111	6374	738	94.7 %	AAC32457 [2e-17]	aminotransferase	aminotran_1 aa 2-385;
5	9590	8433	1158	74.4 %	CAB42045 [9e-67]	L-dopa decarboxylase	pyridoxal deC aa 48-411
6	11393	9855	1539	85.5 %	AAD21754 [e-113]	oxidoreductase	short chain dehydrogenase domain aa 53- 240;
7	13656	12712	945	86 %	CAB41201 [7e-41]	polyketide synthase	ketosyl-synthase aa 13-438; acyl transferase aa 533-854; short chain dehydrogenase domain aa 1159-1357; pp binding aa 1451-1515;
8	15374	19584	4611	87.8 %	CAB06094 [e-180]	polypeptide synthetase	DUF4 aa 17-409; AMP-binding aa 507- 905; pp-binding aa 991-1054; DUF4 aa 1067-1466; DUF4 aa 1544-1844; AMP- binding aa 2041-2439; pp-binding aa 2525-2589;
9	20003	27889	7887	88.0 %	AAC80285 [e-180]	peptidase	Signal70 ECF aa 17-83;
10	28251	29402	1020	86.6 %	BAA13079 [2e-44]	sigma factor	HTH aa 511-109;
11	31720	30401	1320	98.9 %	no match	regulation	response reg aa 46-159; signal aa 326- 542; response reg aa 590-703;
12	31982	32932	951	95.3 %	CAB09733 [2e-63]	regulation	
13	33128	33613	486	95.7 %	no match	regulation	
14	33661	34077	417	94.2 %	CAA19900 [e-37]	regulation	
15	35811	35255	357	80.8 %	CAA19859 [3e-15]	regulation	
16	37856	35730	>2127	88.8 %	BAA17685 [2e-50]	regulation	

## Claims

5

10

15

20

25

30

35

40

45

50

55

5

45

## Claims

10

1. DNA sequence, the expression products of which cause an enzymatic biosynthesis, a mutasynthesis or a partial synthesis of polyketide or heteropolyketide compounds or are involved therein.

15

20

10

2. DNA sequence according to claim 1, wherein the polyketide or heteropolyketide compounds are epothilones.

25

15

3. DNA sequence according to any of the preceding claims, wherein the DNA is derived from myxobacteria.

30

4. DNA sequence according to any of the preceding claims, wherein the DNA is derived from Sorangium strains.

35

20

5. DNA sequence according to any of the preceding claims, wherein the DNA is derived from Sorangium cellulosum.

6. DNA sequence according to any of the preceding claims, wherein the DNA is selected from the group consisting of:

(a) the following DNA sequence:

40

25

Seq ID No 1 (A2 cosmid)

45

30

GGATCGCGGCGCCCTCGCGCTGCTCCTCGAGCGTGCGGAGGAACTCCCACGCCAGGCCCGACT  
TGCCGAGGCCAGGCGCGCCACCACCACCACCGCTTCGCGGAGGGCTCGTCGACGCAATGGC  
GCCACTCGGTGCGGAGCTGCGAGAGCTCGCGCTCCCGCCCCACGCAGGGCGTCGGCTTGCCGA  
GCAGCCGTGGGACGGCATCCGGCTCCTCCTTCGGGCCGCGAAGCCAGCACCTCCGGGCCCT

50

55

5

46

10

15

20

25

30

35

40

45

50

55

GTACCGTCTCGAAGCGGCTCGCGAGCAGGCTGGCCGTCGCGTCGTCGAGCCGGATCTCCGGCG  
GCGACAGGCCATCTCGCCCGCGATGAGCTGCGCGACCCGATCGACCAGCTCGCCGACCGGCA  
GCCTCGCCTCGACCTCGGCCAGCCCTGTGCGGACGGACACGGGCACGCCTCCGAGCGCCGCCC  
GCAGCGCGAGGGCGCAGTGGGCCGCCCGTGTGGCGAGATCCGTGGGCGACTCGGCGCCGGACA  
5 GCGCGACGAGCCACCAGCGCGCTTGACGCCGATCGAGGCGCCCGCTGGCGCGCCCGATGT  
15 CCCGACGCGCTCGGCCCGCGCGGCCCGTCTCTCCGAGAGCGTGGCGCCGGCCTCGGCCG  
CGCCGTCTTCGGCCAGGATGACGCACATCACCTTGCGCTCGGCCGTCTGTATCGCCTCGCCCG  
GCGCGGCCCGCGCGCGACCGCGCTCGCCCGATCGAGAGCCCTCGCCGGCCACGGCGGCGA  
GCTCCGCCCGGGCGGCGCGCGCTCGCGCGGCCGTCTCCCGCTTCTTCGCCAGCATCCGCG  
20 10 CCACCAGGCGCTCGAGCGGCTCCGGGATACCGTCGCGGAGCTCCCGAGCCGCGCGGCTCTT  
CCAGGACGACCCGCATCAGGAGCGCGAGCGCGCTGTTGCCGAGGAACGGCGGGCGCCCCGCGA  
GGCACTGGAACAGCACGCACCCGAGCGGAACACGTCGGCCCGGGCGTCGACCGCGCGTCGC  
25 CGCGCACCTGCTCGGGCGCTATGTACCCGGCGTGCCGAGCACGGCCCCGGGCGACGTGAGGG  
TCGGCGGAGCCGGAGGTGGCGCGCGATGCCGAAGTCGAGCAGCGTGACGCGCTCGACCGCGC  
15 CGCCACGAGCATCAGGTTGCTCGGCTTGAGGTGCGGTTGAACGACGCCGAGCCAGTGGATCG  
CGCCGAGCGTCGTGGCCACGCGCGCGGCCAGCGCCACGCTCTCGGCCAGCGTGAGCGGCGCCC  
30 CGGCGAGCCGCTCCTCCAGGGTCACGCCGTGAGCCACTCCATGGCCAGGTACGGCCGCCCTG  
CGCCGGTCACCCCGTGCGCCACGTACTGCACCACGCCGGGAGCCGAGCGTCACGAGCGCCT  
CCGCCTCCCGCGCGAACC GGCGCAGGTGCTTGGCGCTCGCGCCCTGCAAGACCTTGAGCGCGA  
20 CCGCCTGCCCCGACACCCGGTCGCGCGCCCCGTACACGTCCCCCATCCCGCCGAGACGGCGA  
35 GCGCCTCGATCTCGAAACGATCCTCGATCACATCCGCTGCGCGCATGGCGGTGCCAATGTACT  
CCGCGCGAGCCTCGGGCCCCCGCGCGTAAGTGCGGCCCTGCGCCCGGTTGAACGCCAGCCCGA  
GCGTGACCGCCTCGCGCTCGGGATCCACGGCCGCCGATCGGTCCACGCCTCGACGAGCGCCT  
40 GCGTTGAACAACCCGCCACCGGGCGCACGCAGCCGGCATCGCCGCGCTGGCCACCCGGCGCTG  
25 CCGCCCTTAGGCTCACCTCCGCGATGCCCGCTGGTTCAACACGGCAGGTCCCTGCAACCCGG  
CCGATCACTACATGCTCCCGGCCGAGGAGCGCTTGCCCGAGTGC3CGATCTGGTCGATCGCA  
AGGCCTACTTCGTCTTGACGCCCCGCGGCAGATCGGCAAGACGACCTCGCTGCGCACGCTCG  
45 CCCAGGATCTACGGCCGAAGGGCGCTACGTGGCCGTCTCTGCTCGGCGGAGGTGCGGCCCC  
CCTTCTCTGACGATCCCGGCGCGGCCGAGCTCGCGATGCTCGCAGAATGGCGCGGCACCGCCG  
30 GCGCGCAGCTCCCCCGCGATCTGCGGCCGCCACCGTTCCCCGAT3CGCCCGCCGGTCAGCGCA

5

47

10

15

20

25

30

35

40

45

50

55

TCGGGGCCGCCCTGCGCGCCTGGGCTCAGGCCGCGCGCGCCCGCTCGTCGTCTTCCTCGACG  
AGGCCGACGCCCTGCGCGACGCGACGCTCGTCTCCCTATTGCGCCAGATCCGCAGCGGCTATC  
CCGACCGCCCGGTGACTTCCCGCACGCGCTCGCCCTCGTCGGCCTGCGCGACGTGCGCGACT  
ACAAGGTCGCGTCGGTCGACAGCGGCAGGCTCGGGACGTGAGCCCCCTTCAACATCAAGGTCG  
5 AGTCGCTCAGCTGCGCAACTTCAACCGCGACGAGGTGCGCAACTCTACGCTCAGCACACGG  
15 CCGAGACCGGTCAGGTCTTCCGGCCGGACGCCGTGGACCGCGCCTTCGAGCTCAACCGGGCC  
AGCCGTGGCTCGCCAACGCGCTCGCCCGCCAGCTCGTCGAGGTCTTCGTCAAGGACCGCGCGC  
AACCCATCAGCTCTGCGAACGTGCGATCGCGCCAAGGAAATCCTCATCGAGCGGCAGGACACAC  
ACCTCGACAGCCTGGTGGATCGGCTGCGCGAGCCGCGCATCCGCGCGGTGATCGAGCCGATGC  
20 TCGCCGGCACCGCGTTGCCGAGCGTGCCCCCGACGACCTTCGTTTCGCGATCGACCTCGGCC  
TCGTGCGCATGACCGCGGAGGGCGCCCTCGACGTGCGCAACCCCATCTACCGCGAGATCATCG  
TCCGCGAGCTCGCGTTCCCGATCCCGCCCTCACTCCCCCAGATCAAGGCCACGTGGCTCACGC  
25 AGGACGGCCGCTCGACCGGACCGCCTGCTCGACGCTTCCTCTCCTTCTGGCGCCAGCACG  
GCGAGCCGCTCCTCGCGCCGCGCCCTACCATGAGATCGCCCCGCACCTCGTGGTGTGGCCT  
15 TCCTCCACCGCGTGGTGAACGGCGGTGGCACCGTCGAGCGCGAGTACGCCATCGGCCGGGCA  
GGATGGATCTCTGCGTTCGTTACGCGGGCGAGACGCTCGCGATCGAGCTCAAGGTCTGGCGAG  
30 ACGGCCGCCCCGATCCCGTCGCGGAGGGGCTCGCCCAGCTCGACGAGTACCTGGCCGGCCTGG  
GCCTCGATCGCGGATGGCTCATCTCTTCGACCGAGCGCTCCGGACAGCCCCCATCGCCGAGC  
GCACGCGCCGCGAGCGCGCGCTCTCCCCCGCCGGCCGCGAGGTGCGCGTCATTGCGCGCTGAG  
20 GGAGCTCGCCGCGCGGCGAGCGCCCTCCACGAGGGCCGGCCACCTCGGACAGCGTCTCTACT  
35 CCTCCGAGGCCGCGCGCCCCCGCCCGGCCGCGCGCGCGCGCGCTCCAGCTCGCAGC  
GCACCACCAGGACCTCGCCATCCGCGAGCTCCGGCCGCTCCACGAGCGCGTGCAGCGCCGCGC  
GCACCGCGGTGAGCACGTCTCCAGCGCGCGCTTCAGCCGCGCCAGCGTCGCGGCGTTCCGCC  
40 CGAGCGCGAGGTGCGTACGACGCGCCCCACGCTCGCGCCGAGCTCGCTCTTGCGCTTGTTGA  
25 CCGCCGCCATCGCCCGCCCGCCAGATCCAGAGCCCCGATCCGAAGGCGCCGCGACCGCCG  
CGAAATCCGCCGCTGAAGGCCACTTCGCCCCGTGGATCGAGGTATCGCCCCGTCTCTCCGCGT  
ACACCCAGCGCCAGACCTCGTCGGTGATGTACGGCAGGACCGGCGCGAACAGCCGCGAGCAGCA  
45 CCGACAGCCCCGAGCCGAGCGCCGCCACCGCCGAGCCGCGCGCCGCTCCCCGGCGCCGCCCT  
CGCCGCGCGCCCGCGCCTTCGCGAGCTCCAGGTAGGCGTCGGTGAACAGCGCCAGAAGAAAT  
30 CCTCGGTCCGCTCGAGCGCCGCCGCCAACTCGTGCTCGTCGAACGAGCGCGTCCGCTCGTCCA



5

48

10

15

20

25

30

35

40

45

50

55

CCACGGCCGACAGCTTGTGCAAGAGCGCCCGGTGAGCTCCTCGGAGATCGGGTGGACCTCCG  
CCGACTGGCTGAGCACGTACTTGCTCGCGTTCCAGATCTTCGTGACGAGCCGCTTGCCGATCT  
TCAGCACCTTCTCGTCGAACGCCGTGTCCGTGCCGAGCCGCGCTCGCCGACCAGTAGCGGA  
CCGCGTCCGAAGAATACGTGTCGAGCAGGTGCATCGGCGTGACGACGTTGCCCTTGCTCTTCG  
5 ACATCTTCTTGCGATCCGGATCGAGGATCCACCCGAGATCGCGACGTGGTGCCACGGGACCG  
ACGACTCGTGAGCATCGCCTTCGCGATCGTGTAAGACGCCACGTCCTGATGATGTCTGTTGG  
15 CCTGCGGCCGAGATCGGCCGGGAAGAGCCGCGCTGGCGCGCCGGATCGTCCCCCAGTGAG  
AGCTGATCTGCGCGTGAGCGAGCTCGTGAACCACGTGTGAAGACGTGCGACTCGGCGGTGA  
AGCCGCCGGGCTGGTCCCGCTGCGACGCTCGTACCCGGGCGGCACGTGACCGTGGGTGCGA  
20 CCGGGAGCATCTCGCGCTCGCGAGCAGCGCCGGCTGTGATCCGGGTGCCCTCGGCGTGA  
GCGGATACCAGACCGGGAAGTGCACGCCGAAATACCGCTGGCGGCTGATGCACAGTCAACCT  
GGAGCCCCCTCGGTCCAGTTGCGGTACCGGAGGCGCATGAAATCCGGGTGCCACTTGATCTTGT  
CGCCGTATTTCAGGAGCTCGGCCTTCTTGTCGCGAGCCGACGAACCACTGCCGCTGGGCA  
25 CGAAGTTCGAGCGGCTGGTGCCTCCGCTCGTAGAAGTTCACCGCGCGCTCGATCGGCCTCGGCT  
15 CGCCCCGAGCGCGCCCGCCCGGCGCGCGCGCGCTGCTCCTCGCGGCGGAGCAGCTCGA  
CCACCGCCGCGCGCGCTGCTTACCCCCCTGCCCTGGAGCGGCGCATACGCGGCTTGCCCG  
CGGCCGGGTGCGGCTCTCCACGCGCCCTCGCCGAACGTACCGGCAGGACACGGCCGTTCT  
30 TGCCGAGCATCTGCCGAGCGGGAGCTTCTGCTCCCGCCACCAGATCACGTGGTTCGCTCGC  
CGAAGGTACAGACCATCAGGATGCCCCTGCTTCTCGCGATCCACGAGCGGGCTCGGGAAGA  
20 TCGGCACCGGCGCGCGGAAGATCGGGGTGAGCGCGCTTTCGCGAAGAGGTGCTGATACCGG  
35 GGTCTCCGGGTGCGCGGTGACGCCGACGACGCGGCGGAGCAGCTCCGGGCGCGTCTGGCGA  
TGACGAGCTCCTCGGCCGTCCCTCCACCGGAACGCGATGTCTGGAACGCGCCGATTGCG  
GGCGATCCTCGACCTCCGCCTGGGCGACCGCGGTCTGGAAATCGACGTCCACATCGTGGCG  
CGAAGACCGAGTAGAGGTGGCCCTTCTCGTGGAGATCCAGGAACGACAGCTGCGCCGCTCTGC  
40 GGCAATGATCATCGATGGTGGCGTACTCGTTCCGCCAGTCGACCGAGAGGCCACCCGGCGGA  
25 AGAGCGCCTTGAAGACCTGCTCGTCTCGCGCGTGACCTTGTGGCAGAGCTCGATGAAGTTGG  
GCCGCGACAGATGCGCGGCGGCTCCTTCTTGATCGTCTCCGGCGCGGCTGCGGCAAGGTCA  
45 GGCCGCGCTCGTACGGCGTGCGCACGTGGTGCGGACGTGGAAGTAGTTCTGCACGCGCGCT  
CGGTGGGAGGCGGCTTGTCTCCAGCCCATCGGGTAGAAGATGTTGAAGCCGCGCATCCGGC  
30 GCTGCGGACGACGACGTCCGTGTGCGTGTAGCTGAAGACGTGGCCGATGTGCAGCGAGCCCC

5

49

10

15

20

25

30

35

40

45

50

55

AGGCGGTCGGCGGCGGGGTGTCGACGACGAAGGTCTCCTCGCGGGGCGCGACGGGTCGTATC  
GGTACGTCCCGTCGGCCTCCACAGGTCCGCCAGGCGCAGCTCGGCCACGGGCGAGTCGAAGT  
GCTTCGGGAGCGTCGCGGGATCGATGGAGCGGAACGTCTTCTTGATCGTCACGTGGTCACCTG  
CAGAACAGACCCCGCAGGAACCGCCCGGGGCGCGCATCCTACGTCTGCCCGGGTGCCTG  
5 TCAAGGCGCGCCGCGCCGCGCGCGCGCGCTCCGTGGAGCCGAGAGGAGAGGCCGCGCCAGGTC  
GGCGGCTCCGCGGCGCGCGCGCGCTCCGTGGAGCCGAGAGGAGAGGCCGCGCCAGGTC  
GTGGAGGACGCGGCGGCGCGCGCGCGGAGATCGCGGAGAGGCGGGCGCATCGATCGCGGCGA  
GGCCGGGGGCTCAGTCGTAGCGCTCGACGTGGACGTGCTTGCGGTGGACGCCGAGCTCGCCG  
GGCGGAGCTCGCGGACGGACGAGACCATCCGATCCAGGCCGAGATGAAGACGTGCGGCGCGG  
10 GATCTCCGCTCTTCTCCGCGAGCTCCCGGTAGAGCTCGGGCACGTGCGCTGCACGTAGCCG  
GGCGGCCGGCCACGACGGGCCGCGCGCGAGAGCGTGATCTCGTAGCGGATCCGGTCGGATC  
CGCGCGGAGCGCCTCGAGCTCGTCGCGGTAGATGACGTCTCTCTGAAGCGCGCGCGAACA  
GGATCCACAGGTGGGGCGCGGCCAGCCCGCGCGAGGAGGCGCGCAGCATGCTCCGAGCG  
25 GCGTGATGCCGGTGCCGGTCGCGACGAACAAGGAGGGCGCGGAATCCCCGGGATCGCGGGTGA  
15 AGAGCCCGTGCGGGCCGATGGCGCGGAGCGTGCGCCGGGCTCGAGCCGGTGAGGTGCTCCG  
AGCCCCCCCCGCTGCACGAGCGTGACCGCGAGATCGAAGCGGGGCGAGCCGTGCGGCGCGG  
ATGCGATGGAGTAGGCGCGCTTACCTCGCCGCCCGGAGCGGGAGGACGAGGTTGACCCACT  
GGCCCCCTCGAACAGAAACGACCTCCCGTCGGCGCGCTCGAACGAGAGCTCGCGCACGAAAG  
GGCTGAGGGGCGGGCGGCGACGAGGCGGGCTTCGAACGGTTGCGCGTGATCATGGTCGGGG  
20 CCCGGCGGGGCTCGGCTGCGAGGCCGCGCGGGTGGCGAGGTCTTACCGCAGCCTGCGCCCCG  
35 CCCAATCGCGATCGCCGCGGAAGGGCGCCCGGAGGGCGCGCAATCGCGGGAATCACGGGC  
TTCCGCCCCGTGCGCCGCGGAGCGCGCGGGCGGGCCCGGCCCGCGCTCCGGCGGGGAGCC  
GTCGCGGGCTCTACCGCACGCCCATCGCGCGGCGTGCGGGATGTTACCGCCGGCGGGAGC  
40 GATCCTGGTTGGGGAGCGCGCGCGGGCGGGGATCCCGGTGCGCGGGCTTCTGCGCGGGGA  
25 GCTGCCCTCGCTGAGCCGGGCGCTGGTCGGGCGACTTGCCGAGCCCAGCGCGAGATCGGAGA  
CGGGGAGATGCGCGCGTCGCTGCATAGAATCCTCCATGGAATCGGTCATCAACACATCGGGAA  
GAGCACCCAGGTGAAAGAAACCTTCGAAGAACCGGCTCTCATACACCTCCATTTCATCGTGC  
45 GACCCCGGATTGAGGACGATCGAACCCGCGAGGGACGCTGGCTCTCTGGGCCTCTCCCTGCT  
CGCTCGACCGGCGCCCTCTCGACGCAACTCCGCCGTTCTGTCGGGACGGGACGGTCCGCTCGC  
50 CGCACGCTCCCGTCGAGACGACTCAGCGTCTCGACGTCAGGAGAGATGACGACTCGGCCCGT

5

50

10

15

20

25

30

35

40

45

50

55

CGCGCCACGACCCCTTCCGGCTCGGTGCTTCGAGCGCGCGGCCAGCGAGCGAGGGGCGATCGCC  
AGGAGATCACGAATCTCCCGGCCATCGGCCTCCAGCGCCTCGGGCTCGTTTCGCTCGTCGCCCC  
GCTCCGTCCCCCGCGCGCACGACGCGAGCTCGCGCGGGGAACCGCGGGCCGCTGTCTGTGGCT  
GCTGATGCGCGACGATACAGGGGGACGCCGTGCCTACCTGGGCAACAGGCGCTCATCTTCTA  
5 CCACGGCGAGCACTACGGTGAGTGCTGCCATGAGTAGGCCCTGAGGGTCCGCGCGACGGAGC  
15 GTGGTGTCAGCGAGAGATGCGCATGGTGACGCGGGCTACGCGTCGAGAGGGACACTAGCACT  
CGACCTCGATCCTGCCCAGCACTTTTTGTTCGGGGAGGGCTGCCCTCCCGCTGGCCGCTGGCCG  
CTGGCCGCTCGCCGCTGGCCGCTCGCCGCTGGCCGCTGGCCGCTCGCCGCTGGCCGCTGGCCA  
TGTCGACGTCGAGCTCGAGCAGCCCCGGGCTGACGGACAGACCCCGAGTTTCATCGAGCCGGT  
20 GATGCCGAACCCGCCAAGCGAAAAACGTATCCGTTTCGGCAGGTCGTGGCCTATCATGCAAGC  
TGCTCGATGCGCTGACAGGCTTCTTCGAGATCCTCGTCGGTCTTTGCGAAGCAAAACCGCATG  
AAGCGACTCCCTGCGTCCCTTCAAAGAAGGCGTCGCCTGGCACGCCCCGCCACCCCGGTCTCG  
25 TCGAGCAAGTAGATGGCTCGCTCTCGACCTGTCTCCCGGGTAGGCGAGACACATCCGCCAGC  
ACGTAGTACGTCCCTGCGGCACGCAGGGTGGCAAGCCCGCTTTCTCCAGCGCCCGACAGAAC  
15 CGGTCTCGCTTCCGTTTCATATCCCTGGGCAAGCCCGTGTAACGAGCGAGGAAGGCCCGG  
ATCCCGGCAGCGACTCCATGCTGCAGCGGCGTCGGCGCGCAGACATACAGCAGGTCGCTCATG  
30 GCTCCAATGGCCTTCGCCCACCTGGCATCGGCCACGCTGTAGCCGATCCGCCATCCTGTGATG  
CTGAAGGTCTTCGAGTAGCCGCCTATTCGTGATCGTACGCTCGGACATGCGCGGAAGGGAGGCG  
ACGCTGACGTGCTCACGGCCGTCGAAGATAAAGTACTCGTAAATTTTCGTCCGTGATCACCATG  
20 AGGTTCATGGTGGCAGGCGAGATCGGCGATCTGTTCCAGCTCCATTCGGCCGAACACCTTCCCG  
35 GAAGGATTTCCAGGAGAGTTTACCACGATCGCCTTGGTCTTCGGGGTGATCGCGCGCTCCAGC  
TCGTCCCGCTCGACATTCAGCTCAGGGATCGCGCCGTCACATACCGCGGAACAGCCTCGACG  
GCGAGGATAGCCTGGCGTGATAGGCATAAAACGGCTCGAAGAGCAGCACTTCGTCCCAGGA  
40 TTGAGCAAGGCCATGCAAGTGGCCTGAAAGGCCCTGTGCTCCGGCGCTCACCGTGATGTCA  
25 GTCTCCGGATCCGCCCGGATGCCATTATGGCGAGCCAGCTTCGCCCGGATCGCATGGCGCAGC  
TCCACGATGCCGTCGAAGCGCGAATATGTATTGCACCCCCGATCCATCGCCTCCTTCACCGCT  
TGAAGGATCACCGAAGGAACCTGGGGTATCACAGACGCCCTGGGACATATTGATCCCATGGACC  
45 TTGGCGCACGCCAGGGTCATGGTACCGATATCGGACTGGGCGAGGCGAGCCGCACGATCACTC  
GGTAGACTCTTCATCAGCGTGCTCCTGCTTCTGTTCTGCGGCTCTGCATGGTGTCTTCGGGTG  
30 GGCTTGTCAGCTCGACGCGCCCATGCAGCGGCGCAGCCCTAGCGGCCGAGGTCTGTCCACAC

5

51

10

15

20

25

30

35

40

45

50

55

TTCTTTGATGAAAGCGAGACATTTCGCTTTTCGTGCCCTGTTTGCCCGCAGCCCTCCAGCCCC  
AGGTACGGGCTTGTTCGGCGGGCCAGATCGAGTACTGCTCTTCGCCGTTTACCACGACCTGGCA  
ACGCGTCTTGCTTTTCGTTCGTCCCGATTTCATGATTTTCCTCGCCCTTCGTTCAGCGCTGCGCGAG  
CATGAAACGAATCGCTCATCGGCGCACAGGCGCGCGCCGGCTGCCCGAGGCACTCCCACGCC  
5 TCCCTCACGGCAACCTCATCGCTCCGGATGTTCCCGATGGCGACTCGGATCGTGACCTGCCG  
15 TGGAGACGGGTATGGGACAAAAATACCTGCCCGACTTGTGACCTCGTCCAGCAGCGCCTCG  
TTGAGGCGATCGAGCTCGCGTTTCGATCGACTCTCTCTGCTCGTCCGCCGACCGCATGATG  
CAAGCGAGCGCGGAGGGCTCATGCGAAAGCAGACCGTACTGAACGGCGTCGGCGCGAGGCGC  
TCCCAATCGGGATCGGCGTCCACCCACTGGGCGAGCTGCTGCCCAATCGGAGGTGCTCCCGG  
20 ATCCGGGCCGCCAGCCCTTCATGCGCGAAGTAGCGCACGATCATCCAGAGCTTCAGCGCTCGG  
AAGCGCCGACCGAGCTGGATACCCAGTCCATGTAATTTCGTGACGTGCGCCCTCGGTGCGGAGG  
TATTCGGGCACCAGACTGAACGCGCGCTTCAGTCGGTCGGCGTCACGCACGTAGAGCACGCTG  
CAATCCATGGGGGTGAACAGCCACTTGTGAGGGTTCACTACCAGCGAGTCCGCCCCCTCGCAG  
25 CCCGCGAGCACGTCCCTGTGCTCGGGACGATCGCGGCCATCCCCGCGTAGGCCGCGTCCACG  
15 TGAAGCCATAGCCCGTGCTCCCGGCAAACGCTGACGATGGCGGGGATGGGGTTCGACGCTCGTC  
GTGGACGTCGTGCCACCGTCGCCGCGACGCAGAGGGTTCGGAGGCCGGCCCCGAGGTCTCTCC  
30 ACGACGGCGGCGCGCAGCGCCTCGGSGACCATGCGGAAGGCCGGATCCGTGGGGATCTTCCGC  
ACCCCTCTGCCCCGATGCCGAGGGTGATGGCTGCCTTCTCGATGGATGAGTGCCTGCTCC  
GACGCGTAGAGTCGCATGCGCCGCTGTCCCGCCATGCCCCGAGCCGGATGGTCGGCTCGGCC  
20 GAGTCGCGCGCGGCCCGGATCGCGA.CCATGCTGGCGGTTCGACGCGGTGTCCATGATCGCGCCG  
35 TGCAAGCCGGCGTCGAGATCCAGCATCTGACGCAGCCAGGAGGACGAGCTCTCGAGCTCG  
GTGGCCCGCGCGACGTGCGCCATAGCATCACGTTGACGTTGAGGCACGCCGCGAGCAGCTCG  
CCGAGGATCCCAGGACCAGACGCCGTGTTTCGCGAAATACGCGAAGAATCGCGGATGATTCCAG  
40 TGCCTGATCCCCGGCAGAATGATCTGCTCGAAATCGGTGAGCACGGCGTCCATCGGCTCCGGC  
25 TCGACGGGCGGGTTCGGGGCCAGCCTGCCCTTCACGTCGCCGGGGCGGATCGCGGGAAGACG  
GGGTATCGATCCGGGTGGCCGAGCTAATCGGCCGCCCAATCGATGATTCTCATACCGATCCGG  
CGGAACCTCTCCAGATCCATGTCCCGAGCCGTTCTTTCCGCGGGTCGCTCACGTCAACTCC  
45 TCGCCCTGCCAGGACAGGATCCTCGAGGTCCCCTGGCTCCGGCGGTGGAAAGCGCTCCTTGAA  
CGTGAAGGCCACGGGGTCGGTCCCTAGCGCCGAGGTGCTCGAGCCGATCCTGCCCCCTCGCG  
30 GACGACGGGATGTGCCCGGCCGGGACCCACCACAGCACGAGGTAATGCGGCTCGAGATGCTC

5

52

10

15

20

25

30

35

40

45

50

55

GAACCACCGAGCGCGCTGTCGCAGGAACGCGGCATGATCCGCGGTGTAGGTGAAGGCGAACAG  
GTGCTCGATGGAGGTCCATACCGACAGGGTCACGAGGAGCCGCTGGTCCGGGTACGGACGGAT  
GGACACAGAGTTCCTTCGGCCGTCTGCAGGCGCCACACGAACCCCTCGCTCCGATCGGCCAG  
ATGGTTGATATGGTTCGAGCCCTGGACGAAGCCCTCCATGATCGGATCCTCCAGCGGAGCGCG  
5 AATACATGCGAAGTTGTATTGCGCGATGTGGTGCCGATGCTCCGACATGTCGCTTTCCATCTC  
CAGCTCCCGCTCACCAATCCCAGCGCTGCTCCGGGGAGCTCATCAGGGCAGACGCGACATCGA  
TCCCGAAGCTCCGCCGCATCCCTTCGACGAAGGCGGCCTGGACCGCTTCGGCGACGGATCGGC  
CTGCCTCCGGCAAGACCTCGGAGACAAAGAAGAACCGCCTCGTGGAAGGGACAATCTTGCCCC  
GCTCCGCCCTGGCGCCATACGAAGTGCCCTCGTCACCAGTCCCTCCGCGTCGGCATAACCGACCT  
10 CGCCGGCGCCGACCGCCACGCTGCCGCCTGAGCCGAGCTCCACGAACGCCTCACCGCCTCCCG  
AGATCTCGAGGCGAACGTCCGGGGCAGCCAGATCGCCGAGGTCCCAAGCGCCGACGGGGACGG  
CGAACCGCAGCGACAGGAGGTTGTAAAAATCGACGAATGCGTTGATGTGCGGCAGCTCTCCAC  
CACCGAGGACCCGCTTCGCCAGCGCCTCGATCGAGCTCGGAAATTTCTTGCCAGAGACCCCCA  
CTCGCTTCATCGCCTCGCGCCAGGCAGCCACGTGCGGATGCGACTGGGCGTTTTCTGTGGCCCC  
15 AGCTCCGTCGCAGCTCCTCCTCGACCTTCCGGAGCTCCTCCAGCACGGCCGGCCGCTCTGCGG  
CGTTGTCCAGGCCTTCCCCGTACCCGGTGACCAAGATCATCCAGGAAACGACTCCAGATTCTC  
GCGGATCGACGATGAATGCCATGTGCCTCCTGCCCCCTCGAGAGCGATCGCCTCGATCGACACC  
AGGCTGTGGATGCATGAGCCGGGCGGTGCGGACGCGAGACCCCGCTACTCATGGCTCTTCGTG  
GCCGATGAACAGGTCTCCACCCGTCGATCGTGCTCGGTGCCCCGATCCGTCCAGTCCCACCC  
20 GCCGGCGACCGCGATGTTTGACCCGAGACGTACGAGGCGCGGTTCGGAGGCGAGGAACGCCAC  
AGCATCTGCGACCTCGCTGGCGCGCCCCAGGCGGCCCATGGGGACGCGCCGCTCCATCCACTC  
CTTCTGCGCGGGCGGAAGGTATCCGTTGTGCGATGAGCCCTGGAGACACACAGTTGACCAGGAT  
TCCATGAGGCGCCTCCTCCGTGGCCAGGCTGCGCGTGAGGATGAGCACGCCGGTCTTCGCGAT  
CGAGTACGCCGCCACGTTTCGGCGCGCCGCGGATCGCGTACGTGGGGCTCAACCCGATATTGAT  
25 GATCCGGCCGCTCTTTGCTGGCGCATGCGGCCACGGCCGCGCGACAGAGGTAATGAACGCT  
GCTCAGGTTGCTGTCCATGACGTTGCGCCATTGCTCGTCGGTTCATCGCCGCAAGCGGCTTGAA  
GAAGAAGTCGCCCACGTTATTGACGAGGATGTCGATGGGGCCCAGCTGCGCCTCGACGCTGGA  
GAAGAGCTCCGCGGCCGCGTTGGGGCGGGTGACGTCGGCCTGCACCACCATGGTTCGTCGCC  
GAGCGCGCGGATCTCGGCCGCCGTCTGCTCGGCCGCATCCTTGTTTGAATGGTAATTGACGGC  
30 GACGTCCGCGCCTTGCTCCGCGAGGCGCAGCGCGATCGCCTTGCCAATTCCGCGCGAGCTACC

5

53

10

15

20

25

30

35

40

45

50

55

GGTGACCAGGGCGACGCGCCCGGCGAGCTCCAGCGATCGCGCCTGTGGCAGGGCCGGAGCAGC  
CTCCTGGTGGAGCTCGACGTCGACGGGGAGCTCCACGTGGTAGCTCGTCTCTCGCGGAGCCGC  
GCAGTACCTCTCGTAGAACGCCTCGAGGACGGGCTCTTCGCGTCGCATGATGTCCGCGTGGGA  
TTCGGCGCTGCGCCACGGATAGAGCACCAGGATCTCGTCGGGGCGCACGGTGCTCTGAAAGAA  
5 TCGCGCATGACCGCACCACCCCGGGTGTTCGTGGGCTCTGGGCGGAGCAGATCATCCATTTT  
TTGCATGATCTGCGTGGCCTCGCCCTCCATGCCGGGTTTGATGCGCCATCGCTCCATAACGAG  
GATCATGTCTTGGCTCCTGTTCGTTCATCGCCGTTTCGATCTGGGGGGGCTGCCCGCGCTCTCG  
AGGGCGCGCCCTTGTATTGGCCCGGATGGTCTGGGTAGCGCTCGCGAGCTTTCGCTTGTGG  
GCGGCGTTCAGGCTTGGCCTTGATTGACGAACCGCTCGCAGACGAATGCGTGCGATTTCATAT  
20 GCGGTCGCGAGCGCCACAGGTAGAGACGGTGGCTACGAGCGAGCTTGGGCGGAAGCGCCCTC  
AAGCGGGTGGATGGATTGAGGGTGCGAATCAAGCGGGCGTGCTCTTGAAGAAGAGACCG  
CTGAATCCATCCTGGACGAACGGCGGGGCCATGGTAGGTGCGACGATGTTGTTGTAATCATCA  
TTCGAGAAATCGCCGTAAACCGGAAAGATGCCGCCGTTGCCACCAGAGCTGCGTGGAGATC  
25 GCGAGCGCCTCTCGCGCCCTCGGAACGTACGGTTGGCGAGCGCCTCCGACAGCCACCTGTGC  
GCCATGAGCAGCGCCTGGACCAGCACGAAGAACGCGTGATGCCCAATACCCAGCGGCCAGC  
GCGCGCTCCGGAGCGCCCGCTCCGCGGAGGCGCAGCCAGCTGCGGGATGCCATTCCATGAT  
TTTGGCCTTCGCTTGCCGAGAACTGGTGGAGGATGTCTCGATGGACGCACAGAGATGCCCC  
ATTTCCATGGGTTGCAGGGAAGTACCTTTCAGGCTTTCGCGGATCATTCGGTAATATGCGACG  
ATCACCGCTTCGCGTACGTCTCGAGCGGGTCTGGGGCGGAGACCCGGTGCACATGGAAATAG  
20 GCGTCGTA TCCGCTCCCGGTCCGACAGGCTCCCGGGATGTCCGGATCTGCCGCTTGCGCC  
TGCCATCGCTCGAGGATGGGGACGGCGAGGGGGCGAGATGCTCGGCGAGCACCGCCAGGTCG  
CCCTCCGGCGCATGCGCCAACAAGACCTCGAAGGCCTCGGCTGCGACCTTGGAGGTCTCGCCCC  
ATTCCGACGGCCTCGATGGCTTGGGGCAGCGGTAGATGGATGGTATATTTAGCCATGATTTGC  
40 CCGAAGATTGCCGCTGCGTCGACAGATCTTTCGCGAGCCGGAACGCCATTTCCACTGCTCTGG  
CTCTCAATATTGAATTGAGCCCTGGCGACTGCCATAGGCCCAGTCGCTCGACACAGTGTAACGG  
25 AGCGGCCCCGATGCTTTCTCCTTTTTCAGTCTGACCGAATACTTCTGTTGGGCGCCAAAGAT  
CCCTTGCCGAGACTGTCCGGCGAGATGTCTGTGCGAAGCGTCCGCACGTCCAGCGGGCCCAT  
GCGTTGCTAGAGCATAAAACGGTTCGATGCCTGGTTCGAGAGGGAGACGCGAGGAGCCTCCCTT  
45 TGGGACGGATGAGGAATTTCTGTGACCGAAATGTCCGCAGGAACAGCGGCGCAGAAGCGGCGCA  
30 TCGATGGGGGAACCATGGGTACGAAGACATTGATGATAATGTGACGCAATCGCAATCGTCGC

5

54

10

15

20

25

30

35

40

45

50

55

GATGAGCGGCCGCTTCCCCGGCCGAGAAACGTCGAGGAGCTGTGGCAGAAGCTCCGCGCTGG  
CGTGGAATGCGTCGTACCTTCACAAGGCCGAGGCGCTCGCCGCGGGGTGAGCCGCGAGAT  
GCTCGCGAATCCCAGCTACGTGCGCAGAGGCGCGCCGCTCGACGGCGTGGAGCTCTTCGACGC  
CTCGTTCCTTCGGGTTAGCCCCAGGCGAGGCGAGAGCATGGATCCGCAGCAGCGCATCTTCCT  
5 GGAGGTCGCTGGGAGGCCCTCGAGCGCGCGGTTACGACCCCGATGCCCATTCGGGGCTAT  
CGGCGTCTTCGCGGGCAGCGCCCCGAGCGGCTACCACTCCCTGGCGCAGTCCGACCCGGAGAT  
CCTAGGCGCCCTCGGCCACTACCAACTGACGCTGAACAACGACAAGGATTATCTCACCACACA  
CGCCTCGTACAAGCTCAATCTGCGGGGCGCCGAGCGTGTGCGTGACAGCTCTGCTCGACCTC  
GCTCGTGGCGTGGTCAATGGCCTGCAGAGCCTGCTCAACCACGAGTGGACATGGCGCTCGC  
10 GGGTGGCGTGGGATCCATGCGCATCAGCGGAGGGGCTATCTGTATCAGGAGAACGGCATCTC  
TTCGCCGATGGGCATTGCCGCGCTTCGATGTGGCCGCAAGGGCACCGTGGGCGGCAGTGG  
CATAGGCATCGTCTCTGAAGCGGCTCGCCGACGCGCTCGCCGACGGCGACACGTCACGC  
GGTGATTTCGAGGAGCGCGATCAACAACGACGCTCGAGCAAGATCGGTTACACCGCGCCGAG  
CGTGAGGGGCGAGCCGAGGTGATCGCATGGCCGAGGCGCTCGCCGGCGTGGAGCCGATGA  
15 CATCAGCTACATCGAGGCGCACGGCACGGGGACGCGGCTCGGCGATCCCATCGAGATCGCAGC  
CCTCACGCGCGTGTTCGGGGCAAGACCGCACGAAGGCAGTTCTGCGCCATCGGCTCGCTCAA  
GACCAACCTCGGCCACCTCGATGCCGCCGCGGGCGTGCCTCGCTGATCAAAACGGTCATGGC  
CCTCGAGCACCGCGAGCTGCCCCGAGCCTGCACTTCGAGCGTCCGAATCCGAAGCTCGAGCT  
GGAGAGCAGCCCTTCTACGTCAACACCCGCTCACTCCGTGGCACGCGGCACGAGGTCCGCG  
20 CCGCGCTGGCGTCAGCTCGTTCGGCATCGGCGGCACCAACGCGCACGTGGTCTCGAAGAAGC  
TCCGGCCCCGCTCCGAGCGGCCCTCGCGGCGTTGGCAGCTCTCACCCTCGCGGCTCGCTC  
CGAGCCCGGGCTCGCGGGGCCACGCGCGACATGATCGAGCACCTCGATCGCCACTCCGGCAC  
ATCGATCGCCGATGTACGTACACGACGACGTCGGGGCGCGGGCTGGCCCTTCGGGCGAGC  
GGTCGTGCGGAGAGCGCGCGGATCTCCGCGCGCGCTCGCGAGCGAGGGCTCGCCGCGCTC  
25 GATCTCGTCATGCCAGGCGGCGAGGAGAGGCCCGTCTCTCTGTTCCCGGTCAGGGAGC  
GCAGCACCTCTTCATGGCGCGGGAGCTGTACGAGGTGAGCCGATCTTCGGCGAGTCCCTCGA  
CCGCTGCGCCGAGCTCTGCGCGGCCCGCTCGGCTCGATCTGCGGCAGGTCTCTACCCCGC  
CGAGGGGCGAGCGGACGACGCGGAGCAGGAGCTCGGTAGGACCGCGATCGCCAGCCCGCGCT  
GTTCGCCATCGAGCTCTCGCTCGCCAAGCTGTGGATGGCCTGGGGGATCGTCCCCAGGCGAT  
30 GATCGGCCACAGCGTCGGCGAGTTCCCGCGGCTTGTCTGGCGGCATCTTCGCGAAGAGGA

5

55

10

15

20

25

30

35

40

45

50

55

CGCGCTCCGCCTCGTCGCCGAGCGGGGCCGCTGATGCAACAGATGCCGCCCGGCGCGATGCT  
GGCGGTGCCCCCTCGCGGAGCCCGAGCTCGCCCCCTACCTCAGCGACGACATCTCGCTCGCGGC  
GATCAACGGTCCGGCTCTCTCGGTGGTCGCTGGGCCGATCGAGGCCATCGACGCGCTCGCGGC  
CGAGCTCTTGGAACACGGGCTCTCGTGCCGGCGACTCCACACGCGGCACGCCCTTCCACTCGAA  
5 GATGATGGCCCCCGTCGTTGACGCCTTTACCCGATGCGTGTCCGCGGTGAGCGCCGCCGCC  
15 GTCAGGCCACTTCTCTCGACCCCTGACGGGCGGCTGGATCTCCCCGAAGCAGCGACCATCCC  
CGCATACTGGGCCCGGCAGCTCGTGGAGCCGGTGCGCTTCGCCCAGGCCGTGAGGCAGCTGCT  
GTCCGAGTCGACGTGGCTCTGGCTCGAGCTGGGTCCGGGCCAGACCCTGAGCCCGCTCGTACG  
GCAGCAGGCCCGCGCGGATGGCGGCCAGGTGGTCGTCGCTCGCTGCCGCGCGCAAGGACGC  
20 10 GGGCGCCGACCACCTCGCGGTCACTGAGGCGCTCGGCCGTGTCTGGAGCGCTGGTGGGACGGT  
CGACTGGAAGCGCTTTCAGAGGGCGAGGCGCGGCGGGTGCTGCTACCGACCTACCCCTT  
CGAGCGCAACGATACTGGGCCTCTCCGCGCCACACGAGCGCTCCGCCGAAGCGATAATCAA  
GCCGCTCTCGGAAGAACCACAAACGTCGCGGATTGGTTCTTCTCCTCCCTGCCTGGCGGCGCTC  
25 GGATCCTCCGGTCTCGTTGACGCGCAGGCGGTGACCACGCGGCGCTCTACGTGGCTCGTCTT  
15 CATCGGGGACGAGGGCCTCGGCGCGGCGCTGGTGGAGGGCCTCGCGCGGCGGGGACGAGGT  
CGTCGCGGTGGTCACGGGTGAGAGGTTGAGCAGACGGGCACGAGCGCTACAGATCGATCC  
30 CGCCGGAATGGCGATGTTGCGTCCCTCTTCGCGCGGCTCGAAATCGAAGGGCGCATGCCGGA  
CCGATCGTCCATGCCTTCTGCACGTCGCTGCGGACGGCGCGCATCGAGCGCGGAGCCGC  
GCTGGAGATCGAGCGCAGGCTGGGCTTCGATAGCCTCCTCCTCCTCGCCCAGGTGATCGCCGC  
20 ACAAAGGCATCCGAAGCCGCTGATGCTCGGCGTGATCACGACCCGGGCGCACTCCGTCATCGG  
35 AACCAGATCATCGAGCCCCTGCGCGCTCTGGTGCTCGGCCCTGCGCGCTCATCCCGAAGA  
AATACCCCATGTCTCGTGCCGGAACATCGATATCGATCTCCCGGGCGAAGGCGGGCGCGCGGA  
GATCGCGGCGCGCCTGATCGCCGATCTGGAGCGAGAGTCGCCCCACTCGGTGGTGGCCTACCG  
40 CGGCGGCGCGCGCTGGGTGAGAGCATAGAGCTCACCGATGTCGGCCGGCGGTGAGCTGGCGC  
25 CGCCCCGCGCCTCCGCCAGCGCGGGGCGTACCTCATTACCGCGGCCTGGGGGGCATCGGCCT  
CGTGCTGACAGAGCTCTTGGCCCCGAGAGGCGCACGCGCTGATCCTGGTTGGGCGGACAGG  
CCTGCCAGCGCGGCGAGGGTGGGACGACTGGCTCGCGGCGCACGGCGGGCGACGCGACGAG  
45 CCGAAAGATCCTCCGGATCCGCGCGCTCGAGGAGGCCGGCGCGGAGGTGAAGATCGCCGCGGC  
CGACCTCTCCGATTTCAATGCGATGCGGAGCGTCATCGAGGAGGCCCGACGCGCTTCGGCCG  
30 CATCGACGGCGTCATTCACTCCGCCGGCATCGCGAGTGGAGGCATGATCCAGCTCAGGACGCC



5

56

10

15

20

25

30

35

40

25

45

30

50

55

GATGGCGGCTTGGCGCGTGATGGGCGCGAAGGTCGGCGGCACGCTCGTGCTCGATGCGCTCCT  
CCGGGACGAGCGTCCCGACTTCCTCCTGATCTGCTCGTTCGCTCGCTGGTCCGCGGCGC  
CAGGAGGAGTATGCGCGGCAACGCGCTTCCTCGACGCGCTACGCGCAGAGCCGCGAGGG  
CGAGGAGGATGCGCGCTCATCTCGGTGCAATGGACACGTCGAGTGACGTCGGGATGGCGGT  
GGACTTCAAGCTCCCGGCCGATCTCCAAGAGGGGCGCGGAGAGCCTGAAGCGGGGCATCAG  
CTCGAGCGAGGGCGCGGAGGTGCTCGGCGCATCTTGAGCGCAGGCATGAGCGGCCCCGCTGGC  
GATTTGCACGTCGGATCTACCAGCGTACAAGCAGTCTGTACGACACGCGGATCGCAGCACGA  
GCAAACTCCCGCGCGCGGCCGATGCACTCGCGCCCAACGACCACGGGAGCCTATGTGCTCC  
CGAGACCGAGACCGAACGGCGCATCGCGCGCATCTGGCAGGATCTCCTCGGCCTCGAGCAGGT  
AGGCGCAAACGACGATTTCTCCAGCTGGGCGGCCATTCTGCTGTTGGCCACGAGGTCCTGTG  
TCGCGTCTGCGAGCCCTCAAGGTGGGATCTCGTTGCCGAGTTCCTTCGATGCGCCGACGGT  
CGCAGGGCTTTCGCGCCTGGTCGACGCAGCACGGGCCGAAGGCGCGGACCCGTCGCGCCGGC  
AATCGGCCGTGTCGAGCGAGACGCTACCGAATCAAGCCGCGCGGCCGAACAGGCCGCCCG  
CACCAAGCCGTAACAAGAAGGGGATCGAGTCATGGAACCCGTCGGCGCGGTGGACATGAATCA  
GCCCGCAAAGCAGCAGGAGACCTGCGTCTTCCCGACCTCCTTCGCGCAGCGGCGGCTCTGGTT  
CCTCGACCAGCTCGAGCCGGGAGCGCGCTCTACAACATGCCCGCCTCCTTCGGACGCGCGG  
GCCGTACGACGTCGACTCGCTCGTGCAGCGTGAACGAGATCGTGCGGCGCCACGAGTCGCT  
GCGCACGACCGTCGATGTCATCGATGGCGAACCCGTGCAGGTGATCGCCCCCTCGTGCGCAT  
CGAGGTGCCCCGTGTCGGACCTGAGCGAGATCGACGAGCCGAGCGAGAGCGGAGGCCCGCG  
GCTCATGGCGGAGGAGAGCCGCGCCCCCTTCGATCTCACGCGAGGGCCGCTGCTCCGAGCCAA  
GCTGCTCCGGCTCGGCGAGGCCGATCACGTGCTGATCTTGACGATGCATCATATCGTCTCCGA  
CGGCTGGTCGATGGACGTGCTGTTCAAGGAGCTTCCACGCTCTACGCGCCTTCCACGAGGG  
CCGCCCCGTGCGCGCTCCCGGAGCTGCCGATTCAATACGCCGACTTCGCGGTGTGGCAGCGGGA  
GCTGCTCCAGGGCGAAGTTCTGGAATCGCACCTCGGGTACTGGAGAGAGCACCTCCGCGGCGC  
CCCCACGCTGCTGGAGCTTCCGATGGACCGGCCCGCGCGCGCAGACGTTCCGGGGCTC  
CCAGCGCGCTTCCGACTCCCACTCTCCCTGCAACAGGCGGTGCAGGCGCTCAGCCGGCAGGA  
AGGCGCGACCCCCCTTCATGACGCTGCTGACGGCGTTCAGCGTGCTGCTCTCGGTTATGCGCG  
GCAGAGCGATCTGGTGGTTGGCACGCCCATCGCGAATCGCACCCGAGCAGAGCTGGAGGGGCT  
GATCGGCTTCTTCGTCAACATGCTGGCGCTGCGCATCGACCTCGGGGGCGACCCGAGCTTCCG  
CGAGCTGCTCGGGCGGGTGGGGAGGTGACGTTGGGCGCCTACGCGCACCAGGACCTGCCCTT

5

57

10

15

20

25

30

35

40

45

50

55

CGAACGGCTGGTGGAGGAGCTGTACCAGGGCGGAGCCCCAGCCACAGCCCCCTGTTCAGGT  
GTCTTCACGTTGCAGAACACCCCGATGGATGCGACGAACAGAGCAGACATTGCATCGGGTGG  
CGCGCCGCTGGTGGAAATGAAGGCGGCGAAATTCGATCTGATCCTGGAGCTCTCGGAATCGCC  
GCAAGGGTTGCTCGGCACGTTTCGAGTACAACACCGACCTGTTCGACGCCGGCACCATCGAGCG  
5 GATGGCCGGCCACCTGGAGGTGCTGCTCTCCAGCGCCGTGCGGGCGCCGGATCGACCCATTGC  
15 GGAGCTGCCGCTCATGGGGGCGGAGAGCGAGTCGGGTATTGGTGGAGTGGAACTCCACTGC  
CGCGCTGTATCCCGAGGACCATTGCATGCACGAGCTGTTTCGAGCAGCAAGTGGAGCGGTGCGC  
CGAGGCGACCGCGGTGCTCCTCCAGCAGCAGACGTTGACGTATCGAGAGCTGAACATGCGCGC  
CAATCAGCTCGCGCATCACCTGCGGAGCCTGGGCGTGGGCCCAGAGGTGCGCGTGGGTGTGA  
20 TCTCGAACGGTCAATCGAGACGGTCTGTGGCGATCCTCGGCGTGCTCAAGGCTGGCGGGGCTA  
10 CGTGCCGCTCGATCCGACGTACCCAGCGAGCGCCTCGGGCTCATGATGGCGGACGCAGCGCC  
CTCGGTGCTGCTCACGAGGCGTCTCTCTCGAAGCTGCCGCCCCACGGGGATGCAACGCT  
GGTACAGCTCGACGCGCTGCACGAAGCGCTCTCCAGGCTGCCACACCATAACCCCGGAGCGG  
25 CGTCACCGCCCAGAACCTCGCATAACGTATGTACACTTCGGGCTCGACCGGGCGGCCAAGGG  
15 CGTGCTCGTCGAGCACCGCGGCCTCTGCAACCTGCCACCGTGCAGGCCAAGCTCTATGGAAT  
CGCGCCGGGGCAGAGGCTCCTCCAGTTCGCGCCGCTCTGCTTCGACACATCGTTCTGCGAGAT  
30 CGCGCTCGCGTTGCTCTCGGGAGCGACGCTGGTTCATGGGCACGGCGGACGAGCTTCTCCCGGG  
ACCTCCGCTGGTCGAGCTGCTGAAGAAGCACGCGGTACGGCGATGCTCCTGGCCCCTACCGT  
GCTCGCAGCGCTGCCAGAACACAGAGCGCGGCGTTGCCGCTGCGCGTGCTCACGATGGCCGG  
20 TGAGGCGTGCCCGCGGAGCTCGTCAAGCGCTGGAAGGCACCCGGACGGCGCCTGTTCAACTC  
35 CTATGGCCCGACCGAGACGACCATTGGGGCAAGCTCCGCAGCGGACCTGTCCGACGAACGGAT  
CCCGCCCATCGGCCGTCCGATTGCCAATACGCAAATCTACGTGCTCGACGAAGCGCTCGAGCC  
GGTGCCCATCGGCGTGCCGGGCGAGATCTTCATCGGCGGCGTGGGCGTCCCCGGGGATATCA  
40 CGGGCGTCCGGACCTGACGGCCGAGCGATTTCGTACCCGACCCCTTCGGGCAAACCAAAGGGC  
25 GCGCCTGTATCGGACCGGCGATCGGGCGCGTGGCTGCCGGACGAAACCTCGAGTTTCTCGG  
TCGAAACGACGAGCAGGTGAAGGTCCGCGGTGTCCGCATCGAGCTGGAGGAGATCCGCGCGGC  
GTTGCTCAAGCACCCGGCGGTGCTCAAGCCGTGGCCGTGGTGGCGGAGGACACGCCGGGGGA  
45 CAAGCGGCTCGTCGCGTATGTGCTCGGACGCGGAGGAGCGCGGTGACCGCCGCGAGCTGCG  
CCAGTCCGTGAGCGAGCGATTGCCTGCGACCATGGTGCCATCGTCCTTCGTGGCGCTCGACGC  
30 CTTGCCCCTGACGCCGAATGGCAAGGTGGACCGCGCGCTGCCGGAGCCCAGCAGAGCGC

5

58

10

5

15

20

20

25

15

30

35

20

40

25

45

30

50

55

CGGCGGCGAGGACCACGTCGCGCCCGCAACGCCGTCGAGGAGGAGCTCGCCAGGATCTGGGC  
GAGCGTCCTCCGGCTCGAAAGGGTCGGCGTCCACGACAATTCTTCGAGATCGGCGGCGACTC  
GATCCTGAGCATCCAGATCGTGGTSCGCGCGCAGCAGGCAGGGCTGCGCCTCACCCGCGTCA  
GATGTTCCAGCACCAGACCATCGCCGAGCTTTCGACCGTGGCTAGAGCCGTCGAGGCGGTCCA  
CGTCGAGCAGGACCCGGTGACCGGTCCCGCGCGCTCACGCCGGTGACGCGTGGTGGCTGGA  
GCAGGAGGCGGCCGAGCCGACCACTTCAACCAGTCGATCTTCCTCGAGGTACGCGAGCGGCT  
CGACGAGAGCGCGCTGGAGCAGGCCATCGCGCATCTGATCGACCACCACGACGCGTCCGGTT  
GCGCCTCGCGCGCGACGAACGCGGCGCCACCAGGTCTTCGCCGCGCCGGGAGGCTCGACCCC  
ATTTACGCGCGTCGACCTCGGGGCGCTGCCAGCGCCGAGCAGATCTCCGCCATGGAGAAGGC  
CGCGAGCGAGGCGCAGCGAGCCTCGATCTGGCCGCGGGCCCGGTCTCCGCGCCGTGCTCTT  
CGACCTCGGCGAGGTGCCCCGCAACGCGCTGCTCGTCATCGCCACCATATTGCGGTGACAG  
CGTCTCCTGGCGGATCCTGCTCGACGATCTCTTTGGGGCCTATGAGCAGGCGCGCCGCGCGA  
GGCCGTACGCCTGCCGCCAAGACCAGTCGGTCAAGCGCTGGGCGGAGCTGCTACCGAGCA  
CGCCGGCTCCGAGGCCGTCAAGGCGGAGCTCGGCTACTGGCTCGACTCATCGCGACGAACGGT  
AGCTCCGCTGCCCGTGGATCGACGGGCGGCGAGGACGTGTGGGGCTCGGCGGCCACATCGT  
CGTCTCGCTCACGCCGAGCAGACGGAGCAGCTCTGCGCGAGGTGCCGAGGCGTACCGCAC  
ACGGATCGACGACGCGCTCCTCACTGCGTTTCGCGCAGGCCATCGCTCGGTGGACGGGCTCGCC  
GGCGGTGCTCCTCGACCTCGAGGGTCACGGGCGCGAGGAGCTCGCCGGCGTAGACCTCACGCG  
CACGGTCGGCTGGTTTACGGCCATGTACCCGATCCTACTCCGCGTCGACGCGGCGGATCCGGG  
TGAGGCGCTCAAATCGATCAAGGAGCAGCTCCGCGCCGTGCCAGGCCGCGGGCTCGGCTACGG  
CTTGTTGCGTTACCTTCGGTCCGATACCATCGCCGAGGTCCGCGCGTTGCCGAGGCCGAGCT  
CTGCTTCAACTACCTCGGCCAGCTCGATCAGGCGATCCCCGAGGCTGCACCGTTCCGGCCGGC  
GCGCGAGTATCAAGGCTCGGAGCGCAGCCCCGCGGCCATCGCGCCACCTCATCGAGGTGAA  
CGCGAGCATCGCCAATGGGCGCCTGTACGCCACGTGGACGTACAGCGAGCGCCGCCACGAGCC  
CGAAACCATCGAGCGCGTCGCGGCGAGCTTCGTACGGCGCTCCGCGCGCTCATCGCGCACTG  
CACCTTGCCCGAGGTCCGGCGCAACACGCTTCCGACTTCGACAAGGTGCGCCTGCGCCAGGA  
GACCATCGATGCTCTCGACGCAATCGACGCGGGCCCCGGGCGTCTGCGAGGGGGAGCCGAAT  
CGAAGACGTCTACCCGCTCTCGCCGCTCCAGGAGGGCATCCTGTTCCACACGCTCTACGCCAC  
CGATTACACGGCGTATGTCGAGCAGTTCCACTGGACGCTGGAGGGCGATTTCGACGCGGAGGC  
GTTACCCCGCGCCCTCCAGGACGTCTGTCGCTCGGCATGCCGCCCTGCGCACGTGCTTCGCTG

5

59

10

15

20

25

30

35

40

45

50

GGAGCGCCTCGATGCTCCACTTCAGATCGTCCGCACGGGCGCGGTCTCCCCGTCGAGCACCA  
GGACCTACGCGGCCTCGCCGCGGAGGAGCAGACCGCGCACATCTCCCGTTACGTCGAGGCAGA  
GCGCCAGCGCCGGTTTCGATCTGCGAAAGGCGCCCTCATGCGCGCCGGGCTGCTCCGGCTCCG  
CAAGGACGCCTGGTGCCTCGTCGAGACCATCCACCACCTGATCCTGGACGGCTGGTCGACACA  
5 AATCTTGCTCAAAGAAGTGTTACAGCTCTACGAGGCGCACCGCGGACACCGTGCGCATCTCGC  
15 GCTGGAGCTCGAGCAGCCGCGGCCCTACGGCGATTACATCGGCTGGCTCGCGAAGCAGGACCA  
GGTGCACACCGCGGCCTTCTGGCGSCGCGAGCTCGAGGGCTTCTCCGCGCCGACGCCGCTCGG  
CGTCGACCGCGCTGTGCCGCACGACGACGGCGGCCCGCGGTTTGTTGGCGCCGCATCGCCCT  
CTCGGGCGACGACGCGGCCCGGCTCGCCGCCTTCGCGCGTCAGCATCAGCTCAGCATGAGCAC  
20 GCTGGTGCAAGGCGCGTGCGCGCTGCTCTTGTCACGCTACAGCGGCGATCCCGACGTGCTCTT  
10 CGGTATGACCGTCTCGGSCCGCTCGGCGCCGATTCCCGGTATCGAGCGCATGACCGGCCTCTT  
CATCAACACCATTCGGTGCGCGTGCGCGAGCCTGCCGACGCGTCGGTGCTCGCGTGCGTCAA  
25 GGC3CTCCAGGAGCACGAGGACAGAGTGCTCGAGCACGAGCACAGCCCGCTGGTCGAGGTCCA  
GGCCCATAGCGACGTGCCGCGCGGACCCCGCTCTTCGAGAGCCTCGTCGTGTTTCGAGAACTA  
15 CCCGGTGACAGTCATCTTCGAGGCCCTCCGGTCGAGGGGCCGACGCGCGGAGGAGGCCT  
CCGCATGATCGATGCGCAGTATATCAGTGATCCACCGTATCCGCTGACGGTCGTGCGGCCCTT  
30 CCATGGGACGCTTTATCTCAATATTGGCTACGAGCGCCGCGGTTTCGACGACCAGGCCGTGCA  
ACGATGATCGGGCACGTACGACGCTGCTCCGGGGCTTCGTGACAGAGGCCGAGACGTGCGT  
CCGCGATCTGCCGTTGCTGACGGCCGAGGAGGAGCGCACCCAGCTCCACGCGTGGAATGCCAC  
20 GGCCGCGCCGTATCCCGAGGGCCATTGCATGCACGAGCTGTTTCGAGCAGCAAGTGGAGCGGTC  
35 GCGCGAGGCGACCGCGGTGCTCCTCCAGCAGCAGACGTTGACGTATCGAGAGCTGAACATACG  
CGCCAATCAGCTCGCGCATCACCTGCGGAGCCTCGGCGTGGGCCCAGAAAGTGCGCGTGGGCTT  
GTGTCCTCGAACGGTCGATCGAGACGGTCGTGGCGATCCTCGGCGTGCTCAAGGCAGGCGGGGT  
40 CTACGTGCCGCTCGACCCGACGTACCCAGCGAGCGCCTCGGGCTCATGATGGAGGACGCGGC  
25 GCCCTCGGTGCTGCTCACGACAGCTCGCTCCTCTCGAAGCTGCCGCCCCACGGGGATGCAAC  
GCT33TACAGCTCGACGCGCTGCACGAAGCGCTCTCCAGGCTGCCACACCATACCCCGCGGAG  
CGG33TCACGSCCAGAACCTCGCATACGTATGTACACTTCGGCTCGACCGGGCGGCCCAA  
45 GGGCGTGCTCGTCGAGCACCGCGGCCTGTGCAATCTGCCACCGTGACGGCCAAGCTCTATGC  
AATCGCGCCGAGCGACCGGCTCCTCCAGTTCGCGCCGCTCTGCTTCGACACATCGTTCTGCGA  
50 GATCGCGCTCGCGTTGCTCTCGGGAGCGACGCTGGTGATGGGCACGGCGGACGAGCTCCTCCC

55

5

60

10

15

20

25

30

35

40

45

50

55

GGGACCTCCGCTGGTCGAGCTGCTGAAAAAGCACGCGGTACGGCGATGCTCCTGGCCCCCTTC  
GGTGTCTCGCAGCGCTGCCAGAACAACAGAGCGCGCGTTGCCGCTGCGCGTGCTCGCGATGGC  
CGGCGAGGCGTGCCCGCGGAGCTCGTCAAGCGCTGGAAGGCACCCGGACGGCGCCTGTTCAA  
CTCCTATGGCCCCGACCGAGACCACCATTTGGGCAAGCTCCGCAGCGGACCTGTCCGACGAACG  
5 GATCCCGCCCATCGGCCGTCCGATTGCCAATACGCAAATCTACGTGCTCGACGAAGCGCTCGA  
GCCGGTGCCCATCGGCGTGCCGGGCGAGATCTTCATCGGCGGCGTGGGCGTCCCCGGGGATA  
15 TCACGGGCGGCGGACCTGACGGCCGAGCGATTCTGACCCGACCCCTTCGGGCAAACCAAAGG  
GGCGCGCTGTATCGGACCGGCGATCGGGCGCGTGGCTGCCGACGGCAACCTCGAGTTTCT  
CGGTGAAACGACGAGCAGGTGAAGTCCGCGGTATCCGCATCGAGCTGGAGGAGATCCGCGC  
20 GCGTGTGCTGAAGCACCCGCGGTGCTCAAGCCGTGGCCGTGGTGC GCGAGGACGCGCCGGG  
GGACAAGCGGCTCGTCGCGTATGTCTCGGACGCGGAGGAGCGCGCTGACCGCCGCGGAGCT  
GCGCCAGTCCGTGAGCGAGCGATTGCCCGCGACCATGGTGCCGTCTCTTCGTGGCGCTCGA  
25 CGCCCTGCCCTCACGCCGAACGGCAAGGTGGACCGCCGCGCGTGGCGAGCCCGAGCGGAG  
CGCCGGCGGCGAGGACCACGTGCAACCGCGCAACGCCATCGAGGAGGAGCTCACACGAATCTG  
15 GGCCGACGTACTTGGGGCAAAGCGGGTGGTGTGCACGACAATTCTTCGATCTCGGCGGCCA  
TTCCCTGCTGCTCGTCCGGGTGCATGATCGGCTCGGCCAGCGGTTCGATCGGCCGCCCTCGAT  
30 GGTGCACCTCTTCACTATCCGACCGTGGCGTCTGCTCGCGCGGTTCCTTGGCGAACGGGCGAA  
CGGCAAGCAATCGCCGAGGGAGGCCGCGCGGACGTACGGAGCGCGGCCGGCGCCCTGGA  
GGCGCGGGCGCGGGCGGCGAAGGCCATCCGTGGCCCGACCTGACCCGGGCACCTTCCAAGCC  
20 CCGCCGTTCTCGCACATCCGCCGCTCGAGCGCGCGTCCAGCGCCGCCGTTCCGCCGACGAG  
35 GAGGCGGAGACGACGGTCCAAGGCCTTCGTGGGCTCTTTGCCCGCAATCCGGAAGCTGCGC  
GGCAGTTCTGTCGCCCTGCAATGCTGCCATTGTAGAGCTCCTCCGCTCGCCGCGGCCTCTTTT  
CTTGGCGCCCGTCCGCGATTGACCTCACATCCTGATCCCTTCTTGCGTCTCCAGAAAGTGAT  
40 TGACGGCCAGCGCGCGCTTGAGATCTTCCGGCGCGGGCGATTTCATCGCTCCGGCGCGCCG  
25 TGA CTGTACCTGCGAAGGGATTATAATGAAACATAACATTGGGTGGCTTCTACCCGCCGCC  
TCGCGACGCTTGCTTCTGTCGCCGCTGACGCCGAATCACGGTGAGGATGCGCCCTCCGTGA  
CGTCAGCAGAGAGCGGCGCGCGCGGAGCGCTGACTGCGTCTCGGCGGCGAAGCTCCAGG  
45 CGGCGCTGGACGGCGCGCGCGCGCAAAGGCTCCGGGAGCCGAGCGGCGGTCCAGAGCG  
GGGACTGTGTCTGGCGGGCGCCACGGGCGTCTCGGACCTGGTCTGCGAGCACGCCGACGAAGC  
30 CTGAGATCTCTTTCCGATCGGCAGCATCACCAAGACCTTCGTCTCTACGCTGATACTCATGC

5

61

10

15

20

25

30

35

40

45

50

55

TCCGGGCAGAAAGCCGGTTGTCGCTCGACGACGCGGTGTCGAAGTATGTGAAGGGCATCCCCG  
CCGGCGACCAGATGACGCTGCCCCAGATCCTCGGTACACGAGCGGGCTCTTCGATTACACGT  
ACAGCCCCGGCGCTCGGCCAAATGATCGAGGTGGATCCGACCCGCGCCTTCGCGCCGGCAGAGC  
TCATCGCCCTCGCCACGGCCGAGGCGCCGTATTTGCGCGCGGGCGCGGGTTTTCGCTATTTCGA  
5 ACACCAATTACATCGTGGCCGGCCTGGTGGCCGAGGCGGTGTCGGGCGGGACGCTCGCCGGGC  
TGCTCCGCACGCGCATCTAGACCCGTGGGCTCGCGCACACGTATCTGGACGGCGCCGAGC  
CGCCGGTCCAAGGGCTCATCCGCGGCTACGGCGACTACGGCGCGGGCTTGGTCGACATCACCG  
ACCAGCTGTCGCCCACCGAGGCGTGGGCCGCCGGCGCCCTGGTGTGGAACGTCGATGACCTCA  
ATCGCTTCTTTGCCCTGCTCATCAGCCACGAGCTGCTCTCGTCGACGAGCTTCAGGACATGA  
10 CCACCTGGACCCCGACGATGTGGCCCCACGAGCCCGGATATGGCCTCGGCCTCATCGAGCGCG  
ATTCTGCGCTCGGCTCCCTCAACGGGCACTGCGGAATCATCTGGGGCTTCAATCGGCGTCGT  
ACGGGGTGCCCGGCCGCGGACGCGATCACCGCGCTCATCAACGGAGCGACGGCGACGCGAG  
CGCGGCTCGTCGACGAGCTCGCGAAGGTGCTGAAAGAGCGCTGATCGAGGCGGAATGGGAGCG  
CTTCGGCGGGTGGTGTATGGCGCCCGCGCTCAGAACGCGACGCGCAGCCCCGCGCTCAGCGGG  
15 CCTGCGCCGGGCGACGCGGCCACGGCGCCCGGACCGACGAGGAGCCGCGGACGGCGGGCGCG  
CTCGGCGCGTCTGCTCGCCGACCCCGCGCTTGCCGAACACGTAGAGCGGCAGGCCGACGGCG  
ACCCCGGCCACCCCGCCGAGCGCGGTGGCGATCGCCACCTCGGACGCCTCGGCGCGCGCGGCG  
CTGCTGTGCTGTGGCTCGCGAAGACCAGCACCGCGCGCTGAGGATGGCGGGCGCCGCCAGG  
GTCTGTGAGGACGAGCCCCGAGATCACCATGACCGGGCTGTTCCACTCCGTGCTCCGCTCCTCG  
20 AAGTCGCGGAACGCCGCCCTCGCCGCCGCGAGCTCCAGCTCGATCCGGCGCTGCTCGGCGCGC  
CGCTCGTCCGCGGAGCCGATCTCGTGGACGCGCGGGCCTGGCCCTCCAGCGCCCGATGCGC  
GCCTCGTGC GCGCGGGTCTCTCTCCACGTGGCCCTGGCGGGACCCCCGCCACGGCCGGC  
GCGACAGAGGGCGCCGACGCCGGGTGAGGCGGGCGCCGCGGGCGGCTCCGCGGCCACGGAA  
GGCGCCGCCCGCGGGAGGCGCGGGCGGCTCCGCGGCCACGGAAGGCGCCGCCCGCGGAA  
25 GGCGCGGGCGACTCCGCGGCCACGGAAGGCGCGGCCGCCGCGGGAGGCGGGCGGCTCCGCT  
GCAACGGAGAGCGGGGCGCGCGGCGAGCCAGGCCAGCGCCACGCGACGACACGACGGCGC  
GCCGCAACCGCGCGGGGCGCGGAAGCGGAGGTGGACCTGCTCCATGCGCGCAGCGTCGCC  
40 CTCGACAGGGCCGGGTCAAGGCGCGGAGTCCGAGAGCACGAGACCTCCGCGCCGAGGAAAC  
AGGCGCGCCGGCGGCCCGCGCGGCGGCTCGCCGCTCACCCTCGCGCGGCGGGCGGGCGCC  
30 GCCTCCCTCCCCGGCGGGCGCGCGTGGCGGCCACGCGGAGCAGCTCCTGGAAGTGCCGCT

5

62

10

5

15

20

10

25

15

30

35

40

25

45

30

50

55

CCACCGGGCCGAGGTCGATGCCGTCCATGAACGACGTGAACGCGAAGTACGGCAGCAGCGTCT  
GCCAGGCGCGCAGCCAGCCCGGGAGGTAGCGCGGCCGCTCCAGCACGCCGGCGGCGCGCAGCT  
CCGCCAGCACCGCGATGTCGTGAGCGCGATCCGGCCGCACACGAGCAGCTCCACCTCGTCCG  
GCAGCCCGCCGCGAAGGACCGCGGCCAGGAAGGCGTAGACCTCGAGCAGCCCGGCCAGGTAGC  
AGGCGTCTTGTTGAACGGCGCGCCGCCCTCGACGAGCCCGCCGCGGCACACGCGCTGGGCGT  
CGAAGTAGGCGTCGCGGCGCTCGGCGCCGCGCTCGCGCAGGTGCCGGTACAGGTGAGGAAGC  
TCGCGCCCTGCTCGGCCATGTCCACGAGCCGCACCCGCTCGGCGAGCCGGGTGAGGCGCCGA  
TGGAGAGCGAGCGGCTGTAGAGCTCGSCGAAGATGGCGAGCCCCCTCTGCGTGCCTGGTGC  
GGGGGCCGCCGAGCGCAGGAACGCGCACCGCGGCTGCGCCGCGCCGTTGTGCGCGGTGAGCG  
CGTGCCTCTCGACCTCGTGGTCCACAGCCCCCTCGGCTCCACGCCGCGAAGGTGCGCTCCG  
GCCGATGCGGACCCGGCTCATGCCCAGCACCTTGCGCGTGACGCGCGGTGACGCGTGA  
TCTCGAGGTGAGCCGCGCGCCCGCGGCCACGCGGGCGGCGAGCATGTCCCGAGCGCGC  
CGGCGTCGAGCGGCTCCTCCTCGGATCGCTGGCTCGTCCCAGCCGTGGACGCGAGGCGCT  
CGGTGAGGTGCTCGGCGAGGTGATGTTCTGAGCGAGCCGCCGAAGAACCAGCGAGCGCGCG  
CGCCGTAGAGCTCCTGCGACCGCGCGGAGAACGCGCGGCTGCCCGCGGCTCGAGCAGCTCCG  
CGGCTTGATCTGCGCGCGGACGTTGTCCCGCAGCCAGCCGAGCGCCGCGCGTCCCCGTCGA  
TGGCCCCGAGGAGCTCGCGCAGCTCGGCGACGCGCCGCGCGAGGCCGTGCGGATCGACGCGGT  
ACTCGACCTCGGGGAGGCGGTCTCGCCGGCGGCGAAGAAGCGCTCCTCCACCTCGCGCGGCC  
AGGCGATGTCTCGAGCAGCTTGAGGGCCTTGCCCTCCGCCAGGCGGCCGCCACCCGATCGA  
GCTGCTCCAGCACGGCGCGGTGATGCTCATCGAGCGCAGGATCGCCGAACCGCGAGACGCC  
GGAACCGTCATTCCCTCGACGAGGCAACGATTGCCATGTTCCGTGCTTTTTGGAGCGCCGTC  
GTCGCGCTCGCCTGCGGGCTCCGGCGATCCAGCGCGGTTGCATGCAGCGAGGTTGTTCCGGG  
CTGGCTCGAGAGCGTCTTTGGCCCCACACCCGAGACACGAATGCTCCGCGCCGAGCGCGGTTG  
ACCGTGACCCCGCGGAGAGCCGATGATACGCTCCGGCCGATGTGCGAGAGTGTAGCTCAACT  
CGAAGAACACCGCGCGGCGCTCACCGACACTGCTACCGGATGCTGGGTTGCGTGGTTCGACCC  
CGACGACGCCGTCCAGGAGACGATGCTGCGCGCCTGGCGGAGCCTGGATAAGTTTCAGCGGCG  
CTCGTGCCTGCGCACCTGGCTGTACCGCATCGGACGAACGTCTGCATCGACCTGCGGGCCGA  
CCGCGCGCGCGGGCGCGCCCCATCGAGGAAGGCCCGGTGCGCACGTTGGACGACGCGCTCGA  
GACGCGCCCGCGCACCCACTGGCTGGAGCCCGTCCCCGACGCGCACGCCCTGCCGGCGGACAT  
CGACGCCGCGGAGCGGGCGATGCTCCGCCAGAGCATCCGCTCGGTTGCTGCGCGCGCTCCA

5

63

10

15

20

25

30

35

40

45

50

55

GCACCTGCCGCCGAAGCAGCGCGCCGCGCTGCTGCTCACGGAGGTGCTCGGCTGGTCCGCCGC  
GGAGGTCGCCGACAGCCTCAACACCTCGGTGCGCGGATCAACAGCGCGCTCCAGCGCGCGCG  
GGGACGCTCGCGAGCCGCGATCTCGGGCAGCGCGCCCCCTCGCTGCCGGAGCCGAGTCCGC  
GCTGCTCGACCGCTACGTCAACGCCTTCGAGCGGTACGACGTGACGCGCTCACGGCGCTGCT  
5 GCACCAGGACGCGACCCTGTGATGCCGCCGTTTACCCTGTGGCTCCGCGGCCACGAGTCGAT  
CCGCGCCTGGCTCGTGGGCCCCGGAGCGGGCTGCCGCGGGTCGCGGCTCATCCCGACGGCGGC  
GAGCGGCTCGCCCCGCTTCGCGCAGTATCGCCCGGCGCCGGAGGGCGGCCACCGGGCTGGGC  
GCTCATCGTCCTCGACGTGCGGGGGACCGCATCGTCAGCATGACGTCCTTCCTCGACACCGA  
GACGCTCTTCCCGCGGTTCCGCGCTGCCGCTCGATCTACCGGCGTAGCCGCGGGCGCCCTGCCT  
10 GCCTCGCCGCGGGTGCCCTGCCTGCCTAGCCGCGGGCGCCCGGCTGGCCACGGGCGCCCGGC  
CTGSCACGGGCGCCCGGCCAGCGACGGGGCGACGATTTTTTTCTGAGCGACCGATGAGTCCT  
GACGGGGCCGGGGTCTACGGGGTGAATCCAACACGGAGGCACCCATGACCGTGACCATCGC  
CAGCATCGATCATCGTGACCAGGACCTCATGACGGGCCCCAGGCCAAGGCGCCGGCCCGCGC  
GGCGCGCCCGACGCGGCGCCGTCCAGGCGAGCCGTGTGGGCGGGCCGCGTCTGAGCGGGCT  
15 GGCCACGCTGTTCTGACGTTTCGACGCCGCGGTGAAGGTGCTGAAGCTGTTCCCGCGGAGGC  
GTGACCGCGAGCTCGGGTTCGCGCGCACCTCGTCCCCACCCTCGGCTACCTCCAGATCGC  
TTGCCTCGTGCCCTACCTGATCCCGCGCACCGCGGTGCTCGGCGCGATCCTGTGGACCGGCTA  
CCTGGGCGGCGGATCGCGATCCACGTGCGGGTCGAGAACCCTCTTTCAGCCACAGCTCTT  
CCCCATCTACGTGCGCGCGTTCCTCTGGGCGGGGCTCTGGCTGCGCGACCGCCGCGTGCGCGC  
20 GCTGACCGCGAGCCCGTCTGTCGAGGGCCGATGAGCTTCACGTTTCACGAGAGTCCATCACGG  
TAAAAGGAGAAGCGAGCCATGACCACAAAGAACCCCCGCAAGCTCTTCGTCAACCTGTCCGTC  
CGCGACCTGAAGCGATCGATGGAGTTCTTCAGCAAGCTCGGGTTCGAGTTCAACCCGCGATT  
ACCGACGAGAAGGCCGCTGCATGGTCTGTCAGCGAGGAGGCTATGTGCTGCTCCTCGTGGAG  
TCGTTCTTCAAGACGTTTCATGAAGAAGGAGATCTGCAGCACGAGCACGACACGGAAGGGCTC  
25 TTCGCGCTCTCGTGCAGCAGCCGGGCCGAGGTGACGACATGGTGAAGAAGGCGGTGCGGGCG  
GGCGGGTCGCACGCGATGGATCCCGAGGATCACGGCTTCATGTACGGGTGGAGCTTCTACGAC  
GTGATGCGCCACCACTGGGAGGTGATGTGGATGGATCCCAAGGCGATCCAGCCGTAGCCGACG  
45 GGGCTGGGCGCGCCGCTGGAAGAGCCCCCGTGAGGCGGGGAGGCGGGAGGATCACCGTCTTC  
GTAACCCACAGCGATGCAGTATCCGTGCGGCTTCGTATCGAAGCACGGCTGTTACGGGCGCGT  
30 CAGAGCGCGTCGACGGTGTGCGGAGCCGACGACGCGACACGGGCACGAGCGTGGCTCCGATG



5

64

10

15

20

25

30

35

40

45

50

55

GAGATGAGCCGAGTCTCGCCCATGGTCTCGGGGTCATGAATGGATGAGTAGGGGACTCGCTCC  
TTCGTCACGTCGTGCTCGACGGCGACGGCGAGGCCGAGCTCGAAGTGCACGGGGCCTGGACCG  
AAGATCCAGCTCGCCCCGGCGGAGCCCCGACGAAAAGCGTGTCGCCGTCGACGCCAGGGCCG  
TCGTCCCAGCCGGGCGATCCCACCGCGGTGTAGGTGTGTTTCCCGAAGGAACCCGCGAGCGAG  
5 AGTCGAAGTCCGACCGGCGCTCGCCACGCGACGCCCGCTGTGCGCGCCGACGCCGCCGAAGCTC  
15 TCCCCGAAAGGCTTATCCCCTGTCTCGATGAAGCCACCCACCTCGATGACGCTGATGCGGTAC  
GTGAGCGCGAGATTGAGGTGCACCCAGCGCTGTCCGAGCCCGAGTAGAGGCCGGGCCCCACC  
TGCACGCTGAAATCCATGCTCGGCGCGGATCCGCGCGCAGGAGCGACGCCAGGGGCGCTGCCC  
TCCTGCGCGCGGGCCGTCCCGACGCAAAGAAAGAGGGCTGTGCGGAAGAATCCAAGCGAGATC  
20 10 GATCGAAGTGAGCGCATGTGCGGCCCTGGAGCATCCGCTGTACCAGGTGCGTCTGATTTCATGC  
GGCGCGCCGCCGGCGCGCCGCGCTGGCCTGTCCGACGCGAGATCACGAATCCGCCATCGCT  
CCCCTGGGCCGCCGGCGCTCTGGTTCGCCTGCGGGCGTGCGCCGGCGCTCGTGTGGCCCATG  
25 GCAACCTTGTGCGGTGTGCTCGAACAGCACAGAGAGTATCGCGTCCGCAACAACCGCGCGA  
CCCGGCGAGACGCTCGTGGGGCCCCCTGCCTCCCCACTTCATCATAACGCCATCAGGAGCACT  
15 CGACATTTTCATTTCTTCACCTCCACTGGCTGAGGGCGACGGTGTCTCGTCATCGGCCGGTTGCT  
CTGGCGGTTGCTCTGGCGGGGTTTCTGACGCCCCGAACTAACGCTTCGAGCGCTCCCCCTTGC  
30 TCTCCCGTTTCCTTCAGCTCCTCCAGCAGGTCGTGAGGCGCTCGTAGCTGCCTTCCCAGAAGC  
GGCGGTAGTTGTGAGCCAGCCGCTGGCGTCTCGAGCGGCTTGGCCTCGATCCGACAAGGCC  
TCCGCTGCGCGTCTGCGGGCCGCGGAGATCAGGCCCGCTCGCTCCAGCACCTTGAGGTGCTTGG  
20 20 AGATCGCGGGCTGGCTCATCGGAACGGCTTCGCCAGCTCGGTACCGACGCCTCGCCGACG  
CGAGGCGCGGAGGATCGCTCGCCGTGTGCGATCGGCGAGCGCAGCGAACGTTGCGTCTGAGGC  
GCTCGGACGGGGTCATTGCATAACTCCTTGGTATAAAAACCAAGTTAGTTATACAACCTGGGGC  
CCGGGCGGTCAAGCCTCCAGGCGATGGCGGTTGCGCCCGGGGGCTCCGCTCGCGGCACGCGCG  
40 CCGCGCGGCTACGTGCGCGGCGCGGTGAGCACGTCTGCGAGCGTGGCGCCGACCACGGGCTTG  
25 25 GTCAGGTGCAGGTGGAAGCCGGCCCGCTGGACCTGGCCTGATCGTCCGGGCCGCGTAGCCC  
GAGAGCGCCACCAGGTAGAGCGCTTCGCCCGCGGGCGCGGCCCGCGCCGCGCGGCGACCTCA  
TAACCGTCGATGCCGGGCAAGCCGATGTCCACGAAGGCCACCTCGGGGCGCAGCTCCAGAAGC  
45 TTCTTCACGCCCTCCAGCCCGTCCACCGCCACCGTCACCTCGTGCCCCAGCGCCTCGATGTAC  
GCCCCGATCACCCGGCGCACGTCTCCGCGTCTCCACGACGAGCACCCGGCGCGGTCAGCC  
30 GCCGCTCGGGCGCCTCGGCGCGCTGCGCCGAGGCGGCGGGCTCGTCTGCGCTGCGCCGGA

5

65

10

15

20

25

30

35

40

45

GGCGGCCCCCTCGCGGGCGGGGGCGCCCCGGCGCTCGGGGCAGGCTGCGGCGCCGCCCGGGG  
CCGAGCGGCAGGCGCACGGTGAACCTCGCTGCCCTGCCCCGGCCCGGCGCTCGCCGCGGCCACG  
CTGCCGCCGTGCAGTTCCAGGAGCCCGCGACCAGCGTGAGCCCGAGCCCCAGCCCGCCCGTG  
CTCCGGTTCGATGGTCTGGTCGACCTCGGTGAACAGATCGAACACCTTCTCGAGCATCGCCGCC  
5 GGGATGCCGCGGCCCGTGTGCGCGACCCGCGAGCACGGCCTCGGGCGCGCCGACCGCCGCCTCG  
CGCGTGAGGCGCACCGAGATCGAGCCCCCGCGGGGTGTACTTCGCGGCGTTGGTCAGGAGG  
15 TTCGTACCACTTGCTCCAGCCGCTCGCGTCGGCCCCGATGCCGAAGTCCCCGGGCCCCACC  
GACAGCGACAGTCATGGCGCCGGGCTCGACGGCCGGCCTACCCGCGCGCGCGGCTCTGC  
ACCACCGCCGCGAGATCGAGCTCTCGAGGCGCAGCTCCACCGTGCCCCGCGTGATGCGCGAC  
20 ACGTCGAGCAGATCGTCGACCAGCCGCGAGGTGGCCCATCTGCCGCCGCGCGATCTCCCGG  
TAGCGCGCCGACGCGGGCCCGTCGCGCTCCGCGTCGTCGAGCAGCGTCAGCGACAGGCTGATC  
GAGGCCATCGGGTTCCGGAGCTCGTSCGCGAGCATCGCGAGGAACCTCGTCTTGCGCTGATCG  
25 GCGAGCTTCAGCGCCTCGACGAGCGCCTCCACGCGCCTCCGGGCGCGCACCTGGTCGGTCACG  
TCGAACGCGAACACGAAGACGCCCTCGACCGCCCCGTCGCGATCGCGCATCGGCTGGTAGACG  
15 AAGTTGAAGAACACCTCCTCCGTCTGCGCTCGCCCCGGCGATCGAGCCGACCGGGAGCTCC  
TTGCCGACGATGGGCTCGCCGGTGCGGACCAACCGCGTCGAGGAGTCCCAGATGCCCTGTCCC  
30 TCGAGCTCGGGGAGGGCGGCCCGGATGGGCTCGCCACGAGCGATCGACCGCCGACGAGCCGC  
TGGTAGAGCGGGTTGACCACCTCGAAGACGTGCTCCGGCCCGCGGAGGATGGCGATGGGCCCC  
GGGGCTGCGATGAAGAGGTCTTCAGGTACTGGCGCTGCCCCCTCGGCCTCGCGCCGGCGGCGC  
20 GCGAGCTCGACGTGGATGCGGACCCGCGCGAGGAGTCTCTCGCGGAGAACGGCTTCACGAGG  
35 AAGTCGTGCGCGCCGGCCTCGAGGCTGTGACGCGCGCCTCCTCGCCCGCGCGCGGAGAGC  
ATCACACGGCGACGCCGCGGGTGCGATCGTCGGCGCGCAGCGCCTGAGCAGGCCGAAGCCG  
TCGAGCCGCGGCATCATACGTGCGTGAGCACGAGATCCGGCGGGTGGGCGGGGCGCGCTCC  
40 AGGGCGGCCCGACCGTCGGCCACGSCCTCCACCGTCCACCCCTCCGCCACGAGCAGCCGAGC  
25 GCGTACTCGCGCATGTCCGCGTTGTCGTGCGCGACGAGGACGCGCCCCGGCAGCCTCCCGGCC  
GGCCCCCTCGCCCCCGGGCCGGACCCCGGCGCCTGCTCGCCGCGGAGCCACTGCGCGGCCTCG  
TCGAGGAAGGGCGCGGCGTCCCGCCCCCGCGGCGCGCGCGAGGCCGGCGCGAC

or its complementary strand,

30

50

55

5

66

10

(b) DNA-sequences which hybridise under stringent conditions to regions of DNA-sequences according to (a) encoding proteins or to fragments of said DNA-sequences,

15

5 (c) DNA-sequences which hybridise to the DNA-sequences according to (a) and (b) because of a degeneration of the genetic code,

20

(d) allele variations and mutants resulting by substitution, insertion or deletion of nucleotides or inversion of nucleotide segments of DNA-sequences according to (a) to (c), wherein the variations and mutants offer isofunctional expression products.

25

7. DNA sequence according to claims 1 to 5, wherein the DNA is selected from the group consisting of

30

(a) the following DNA Sequence:

Seq ID No 2 (>pEPOcos6 region)

35

20 GGATCACCTGCGGCGGATCGCCGACCTCGTGCTGGTGTTCGGCTCGCTGGATGAGAAGCCGG  
CGGCGCTACTGATAGAGACGGCGACGCCCCGGGCTGCGGTGGAGCGGTTGCGGGAGATGCTCG  
GCTTTCGGGCGGCCACCTGGCGAAGCTGTCTTCGACGGTTGCGAGGTCCCCGAGGCTCAGC  
TGATTGGCCGCGCCCGGCTTTGCGCTGATGTATCTGGCCCCCTACGCCCTGGATTTCGGTTCGGG  
TCAGCGTCGCCTGGGCCTGCCTGGGCATGATCCGCGCTTGCTGGAGACCTGCGCACAGCACA  
40 25 TCCTCACCCGCCGCACCTTCGGCCACCTGCTAGCCGATCACGGCATGATCCAAACCTGATCA  
CCAACCTGGGGATTACCAACAGGCGACGCTGCTCCACACGCTGCAGGCCTGCCGCGCCAGGG  
ATCGCGGCGACGTGACCGCTCCGAGGCCACCCTCGCCGCCAAATACCTCGCGTCGCGGACGG  
45 CGGTCCAGGAGACGACCAACGCGGTCCAGATCATGGGCGCGCTGGGCTGCGACGAGGAGGGCG  
CGATCGCCCGCCACTTCCGCGACGCCAAGACGACCGAAATCATCGAAGGCAGCAACCAGATCA  
30 TCGAGGCGCTGCTGGCCAAGAACATCGCCCGCGCCGGTCGCGACAACATATCGCCGCTTCCTCG

50

55

5

67

10

15

20

25

30

35

40

45

50

55

ATGCGGAAGTCGAGCCCGGTGCGGCCGGAGGCGCACCATGACGAGCGCGGTCCCGACGCGTCA  
AACCAGCCTGCTCGACGACTTCGAGCGCGTCCCGACGTCGATCCAGAGCGGATCGCCGTCCA  
CGCGAGCGAGACGAGCCTGCGCTATGGCGACATGAATGCGCGCGCCAACCGCATTGCCACGG  
GCTACGGGCGCGCGGGATCGGGCCCAATCAAATCGTGCGGTGGCGATGGCCCGACGCCCCGA  
GCTGATGATCGTGCTGTACGGCATCCTCAAGCGCGCGCGCTACATGCCCATCGCCCGCGA  
CGCGCCGCGCTGCGCCGCGATCATATGCTGCGCGAGAGCCAGGCTGCTCTGATGATCGCCGA  
CGAAGAGATCGCGGGACTCGCGGCCCGGTGCTGACGCGCGCGGACCCGTTCTTCGCGGCCAT  
GCCGGACCACAACCCCGAGCCGCGTCACGACCCGACCGACCTGATTTACGTCATCTACACCTC  
GGGCTCGACCGGCCAGCCCAAGGGCGTGCCATGGAGCACCGCGCCGTGTGGAATCGCCTGAC  
TTGGATGCAGGCCAGTATCCAATCGACACGACGAGACGTGATCCTCCAAAAGACGCCGATCGT  
CTTCGACGTGTGGTCTGGGAGCTGTTCTGGTGCCGCTGGCCGGCGCCTCGGTGGCCCTGCT  
GCCGCAATCCATGGAGAAGTTCCCTGGGCGATATCGGCGACGGTGGCGCGGTGCGGGTGAC  
GGTGTATGCATTTCTGTACCATCGATGCTGATGGCCTTCCTTCAGGTGGTGGCGGGCGGCCGA  
GATGGCGGACAGATGAAGGGCCTGCGCTACGTCTTCTGCAGCGGCGAGGCCCTGGCGCGGGC  
CCACGTGTCAGCCTTTCAGGAGCACATCAACCGAGCGGGCAGCATCAGCTTGACCAACCTCTA  
TGGACCCACCGAGGCGCGGTGACGTCAGCTACTTCGACTGCCCGCCCGCGCGTCACTCGC  
GCGGTGCCGATCGACGAGCGATCACCGGCATCCAGCTGCTGGTCATGCGCGACGGCGTGCC  
TCAGCCGCGCGGTGAGGCTCGCCATCGGCGCGCTTGGTTTGGCGCGCGGTACAT  
CTCACGGCCAGACCTGACCGCCGACCGGTTCTGTCCGCATCCAGGCGGCGACGGCCAGCGGCT  
CTACCGCACCGGCGATCTGGTGCGCAGGGACGCGGACGGCGAGCTGGTCTTCCTGGGGCGCAT  
CGACCATCAGGTGAAAATTCGCGTCTGCGCATCGAGCCCGGGGAAATCGAGGCCAGATCAG  
CGCCCATCCCGATGTGGCCGACTGCGCGCTGATTATCGAGCAGGACTCGGAAACCTGCCCAA  
GCTGACCGCTACATTGTCTGTGGCGGACCGGGCTTGACCCGGAAGGCGCTGCTACAGTTCTT  
GGGCGCGCGGTGCCCGACTACATGCTCCCGAACCGCTTCTGACCTCACGGAGCTGCCCGT  
GACCGCCAACGGTAAGCGGACTGGCGCGCGTGTCTCGGCCCGCTCGAGACCCTGCCTCTCCC  
TTTCTCCTGAATCCAACCAATACGAGGGATTATGTTACACCGATTCCACCGACCGTTTCG  
CCCTGAGCCGACCGCTCTTTGCGGGTACCTCGCGCACGATCCGATCGTGAGGGCGTGCTGG  
CGGCGGACCATCCAGGCTGGGTCCTGGTGACCGCGAGCCCGAGCCGCGACGGCGCTGCTGT  
GGGCTTTTCCGATCGGCTCTTCTGCGTGGGCGCAGCTGACACGCTGACCCCGCACGCGCTGG  
CCGAGCTGTTCCACGACCGACTGATCCCCAGGCCGTAAGATCGGGCAGCCGTTTTTCAGG

5

68

10

15

20

25

30

35

40

45

50

55

TTCAGGGCGAGACGGTCGACACCTGGTCCGACCACCTGCATCAGGTGTCGCCGCACGCGACAG  
TCTCCTTCCGCCAGGCATTCCGCTTCGACCGCGACCTCTTCGAGCGGCTGCCAACCAAGCCGG  
AGCTGGCAGAGGCGCGGCTCGTGCCAATCGACGCGCGGCTGCTGGCCGAACAGGCTGATCTGC  
GCGAGCGGATACTGGCCTCCTGGTCCAGCGAAGCTGCCTTCCATGCGCGCGGTTTCGGCTTCT  
5 GCTACCGCGTAGGTGACCAGCTGCCGAGCGTGTGCCTGGCATCGCACGTAGCGGCGGCGCGG  
15 CCGAGCTGAGCATCAACACCGAGCTCGAAGCGCGCAATCGAGGTATGGCAACGCGGCTGTGCC  
GGCGTTTCATCGCCGAATCGCTGCAGCGCGGCTGACGCCTTGCTGGGGCACCGAGACCTTTC  
GCCTGCCGTCAATCGCGCTGGCCCAAGCTCGGTTTCATCCCGACCTTCACCTTCCCCACCT  
ACTGCTTCGCGACCGGCACCGAACAGCCGACGACAACCTTCCTAGGCGAGCTGTACTACAGGG  
20 10 AATCGCGCATCGCCGGAAGTGGGACCGATGAGCCGCAAGCGGTTTCGGCTGGCGCGGGGTGGA  
GCCTGGCCGCGGACACCGAGCGTGCCGCGAGCTTCGCCCGACGCGCCCTGGCCGAAGGGTGGG  
CCGGCCACTCGACTCTGGCCACCGATCCGGATTTGCCCCGATTGCGCGCCAGCGCCGCTGGC  
CCCGCTCAATGTCCCTTGAAAGGTCACGTGGACTCATGATGTCCCTTGAAAGGTCACACTC  
25 CGAGTCATGATGATTTGTCACTCCCACCGCTTCATTTTCTCCACGTTCCCAAGGTCGCCGGC  
15 ACAAGCGTCAAGGACGTCTCGGCCAAGAGCTATTCCAGGAGGACCAGGTCACGTTCCAGATC  
GCTCCCAATCCCCACTACCCACCTGAATGGACTGCGCCTTACGAGGAGCACATTATTGCCGCT  
30 GAATTGAAGAGCCAGTTGGCGCCGAAATTTGGGACGATTACTTCAAGTTCGCCTTCGTGCGC  
CATCCGCTCGACTGGGCGGTCTCCAATTACTTCTTCTTCTGCGCGACCGCAAAGGCCATCCG  
GCCCACGAATTCTTGAGCGGAAGGGCTTCGCCGGTACCATGGACATGTTTTTCGGAGCGGCC  
20 GGGCGCCATCCGCTGGTCGCCGGCATGCGCTTCAGCCAATGGGAGTTCTTGTGCGACAGCGAG  
35 GGCCGGACGCTGGTGGACTTCGTTGGCAAGTACGAGCGGCTCGAGCAGGACTTCGCCGCCGTG  
TGTATCCGCATCGGGCTGACCCCGCCGACTTGCCGTGCCTCAACCAGACTCGCCACCAATCC  
TTTACCAGTTACTACGACGAGGCTTTGATGCGCCAAGTCAGCCGCGGTTAGCTCCGGATTTTC  
40 GAAATTTTGTATTATGCCTGAGGCGGACCCGTTGCTTCGCCACCGGTGGATTATTCGATAAGT  
25 TATTATATTTTCAGTTGATCATGTGAATGTGATCCAGCCAACGAGGAGGATACCTCCGCGTG  
CGGCTATGGGGGCGCAGAGGTCACCACTACGTGTAGAAATTTGTGCAACACACCACTAGCTGC  
CACCGATTGGGAGCTTTGACTTGAAGATGAAAGTGGACAAGCGGAATGTCGACGACATTCTCG  
45 GACTCACTCCGACACAGACAGGCATCTTGTACCACTACCTGCTGGACCCGAGGCCGACGCCT  
ATTTGAAACAATTGACGCTGCACCTGGAGGGGCGCTCGACGTAGCGCGCTTCGCCGCGCCT  
30 GGGAGCGCGTGGTGGCGGCTCACGACCAGCTGCGCGCGTGTTCGCTGGCAAGGGATCGAAC

5

69

10

15

20

25

30

35

40

45

50

55

ACCCGGTGCAGATCATCCTCAAGCAGCACGTGCCGGACCTGGAGTTGGCGGAGGTCCCGCGCG  
ACGCCGATCCGGCAGCCTTCTCTGGC3CAATGGGTTCGCGGCCGACCGGGCGCGCAAGTTCGACT  
TCGAGACGGTGCCCTTTTCGCATCGGCCCTCTGCCGACTGATACCCAACATCACGTGATGCTGC  
TCAGCAATCACCATATCCTGATGGACGGTTGGAGTACGGGCCTGATTCTGCGGGACTTCCTCG  
5 CCTGCTACGGCGACTCCGAAAACCTGGCGGCCACGCACCCGAACGCACTTCAAGGCGTTCATCA  
AGTGGCACCAGAACCGGCCACGCCGGGCGAGGAGCGATTTTGGCGCGACCTGTTGCGCGATG  
CGCCCGACGGCGGCTTTCCCGGCCTGGCGCTCGAAGAAGGCACCCGCCACTCGCTTGACTTCG  
GCGCCCGCAGCCGCGCTCTCGACGACCGCTTGACCCAAGGCTTGCGCGACATGGCTCGCGACC  
TCGACGTCACCCCTCGCCGCGATGCTCCATACCGCTTGGGGCCTTCTACTCCAGCGCTACCAGA  
10 ACAGCTGCGAAGTGATATTCGGGACACCGCTTTCGGCCGCAACGTCGAGCTCGCCGGCCTCG  
ACGAGGTGGTTCGGCTTGTTTCATCAACACGATTCCGTTCCGCTTCTCGGCCGCGGCCGCGACGA  
CGCCCGTCGAGGCCTTCCGTGCGGTACAGCGCAATCTGCTGGCGAGAAGCGAGTTCGAAGCCA  
CCCCGCTGGTGGACATCAAGGGCTGAGTGGTCTCGGTCCGGGCGCGGAACTGTTTCGACACCA  
TCCTGGTCATCGAGAACTATCCCTTGGACCGCGCTATCTTCGAGAGTGATTCCAGCCTGCGGT  
15 TGACCGACCACCAAATCTTCGAGCGACCAATTACGGGTGACCCTGACCATCGAGACCTTCA  
GCCGGTTGCACGTGACGCTAGCCCATCGCCGTGACCTGCTGGGCGACGCGGCCGCTGAGCGAA  
TGCTAGATCATTTACCGGCCTGCTCCAAGCCATGCTGCGCTTCCCTCACCAGCCGTTTCGCGC  
GCCTCGAGATGAAAAGCGAACACGAGGCCACCCGCTCCTGCACCAACTCAACCAAACGCGTC  
AGCCGCTGCCGTCCCAATCGGCTTTCCACCAGTTGTTCTTCGAGCAGGCCAGGCCGATGGGG  
20 CACGACCGGCGCTGTGGTGGCGGCCACGCGCTGGACCTACGCCAGCTGCTGGAACGTGCCC  
TGCGTCTGGCGGGACGGCTGCAGGAAGCCGGCTTCGCCCAGGCGATGTGCGCCCGCTCAGCC  
TCGGCCCGGTTCCGGATCTGATTCCCGGTTTGCTGGGCCCCGCTGTTTCGCCGGCGCGCCTACC  
TGCCGCTCGATCCCAACCTGCCGGCCAGCGCTCGCGGTTTCATCCTCGACGATGCCGGTTGCC  
GCTTCTGATCAGCGACGCGCACTCGCGGGGCCACGCCGATCCATCCGGACCTGCCGGCG  
40 CCAGCCCCGTTGACGTCAATTTTCCCTGTGAGGACGGCGCCGCGCAGCCCGCTACCTGATCT  
ACACCTCGGGCTCCACCGGCCAGCCCAAAGGCGTCTGGGTTAGCCACCGCAACCTGATCAACT  
TCCTGACGGGCATGAGCGCAATCCTGCCGGTCGCGGCCGACGACGTGTTCTCTCGCTGACTA  
45 CCGTGTGCTTCGACATTTTCGGGCTCGAGACGTGGTTCCCGCTCAGCCGCGGCTGCACGATCG  
TCTTGGGCACGCGCGCCGAGCAGTTGGACCCGGCCGCGGCTGCCAAGGCCATCTCCTGCCATG  
30 GCGTCACGGTTTACCAGGCGACGCCATCGCGACTCCAACCTCAACTGGAGCACCCACATTTG

5

70

10

15

20

25

30

35

40

45

50

55

TCCGCGCCATCGGCTCCCTGACGACCCTGCTGGTAGGCGGCGAACCCCTCCCAGCCGAGCTGC  
TCCGGCGCGTACGCGAAGTGACCGATGCGCGTATCTTCAACCTCTACGGTCCCACCGAAACCA  
CCATCTGGTCCACAGCCGGGAGGTCAACGCGGCGGACGTCCCGGATATCGGCCGCCGATCG  
CAAATACCGCGTTTTCCTTCTGGCGGAGACGGCTCGATCCAGCCGCCGGGCTGGTGGGCG  
5 AGTTGTGCATCGCCGGCGAGGGCGTGGCGTTGGGCTACCACCGACGGCCGGACCTGAACCGAG  
AACGGTTTCGCGAGATTCCGCCGGGCCCGCTGCCCTTTGCCGGCAAGCTCTACCACACCGGCG  
ACCTGGCCCGCTGGACCGAAGACGGACGGCTCCTCTGCCTGGGCCGTCTGGACGACCAGCTCA  
AAGTGC GCGGCCATCGCGTCGAGCCGGGCGAGATCGAGGCAGTGATGGCGCGCCACCCGCGG  
TCACGCAGGCGGTGGTCTCACGCGGCCGCGCAACGGCGAGCCGGTCTTGGTCGGGTTCTGGA  
10 CTGCGGAAGGTGAGCCGATGCCAGAGGAAGCGCTGAGCGCTTACCTGGCCGACCGACTGCCGA  
GCTACATGGTACCCGAACGGTGCACTCCTCATGAAGGCCATGCCGCTAACCGGCAACGGCAAGA  
TCGACCGGCGCGCCCTACCCAATCCCTTCGCCTTGACCGAGTCGACCCGGCAGGCGCGCCGC  
GCACCTTGGCCCGCACCGCCGGCGAGCATCGGGTTGCCGAGCTGTGGCAGGCCTTGTTCGAC  
25 GCGAGGCGATCGGCTTGGACGAACCTTTTTCAGGCCGGCGGGAACCTCATTTCGGCTTGATT  
GGCTTCACGCCAAGCTGGAATCCGCCTTCGGGAAGTCGTTCGGATCACCGATTTGTTCCAGC  
ATACAGTATTTCGAGCCAGGCAGAAATGCTGAGCGGCTCGTCCGTGAGGCGCCGCTCGCGG  
30 GAGCCGTGCCGCAACCCCCGGCCGCCGCCGCCAAGTTGCCTCCTCGGCAGCTAAATCCCCAG  
GGGAGCGCGGCGCGGAGCGACGTCGAGCGGCTGACCGCGCAACCGCCCAACCCACTTCC  
GGCCCATCGCCGTTATCGGCCTCGCCGGCCGATTCCCCGCCGACCCGACCTCGACGCCTTCC  
20 TTGAACTGCTCACGGAGGGTGCCTGCGGCATTTCGCTTCTTACGCCAAGCCGAGCTGCGCGACG  
AGGGTCTCGACGCGAATCGAATCGCGTGTCTAACTATGTCCCGGCCAAAGGTTTCTTCGACC  
GGGCCGACCACTTTGATGCCGACTTCTTCGGCATCCCGCCGCGGACGCGAGAAATCACCGATC  
CGCAAATTTCGGCTTCTGCTTGAAGTGTGCTGGAACGCGCTGGAGCATGCCGGCTACCCGCCG  
40 GCGGCGGCGAGATCGGGCTTCTCGCCGGCTCCTCGGCCAACTATCACTGGCTCGAATACGTGG  
25 GCATTTCCGAGGAGAGCAGCAATCGATTGCCGTCATGATTCAAAACGAAAAGGACTACCTGG  
CCACGCGGATCGCCTACCAGCTCGATTGAAGGGCATTGCCGTACCGTGCAAACGGCCTGCT  
CGTCGTGCTGACCGCGGTGAGCTGGCCTGCGATGCGTTACACGCCGGCCGCGTGACCATGG  
45 CTTTGGCTGGTGGCGTTGGTCTGACCTATCCGTTGCGCGCCGATACCTGCACGAGGATGGAA  
TGATCTTCTCCCCGACGCTCGGTGCCGGGCTTCGACGCCAGGCGGCGGCGACGGTCTGCG  
30 GCAACGGTCTGGGCATGGTGGTGTGAAACAGCTCGACGCGGCGCTGGCCGACGCGCATGCCA

5

71

10

15

20

25

30

35

40

45

50

55

TCCACGCTGTGATTAAGGGCATCGCGGCCAACACGACGGCGCGGCCAAGATCGGCTACACGG  
CGCCCTCGCAGAACGGTCAGGCGCGGGTGATCCGCGCCGCCCATAGGCTCGCCCAAGTCGCGC  
CGGAGACCATCGGCTATGTAGAAGCCCACGGTTCGGGCACGCCGCTGGGCGATCCGATCGAGG  
TGGCGGGCCTGACCGAGGCCTTTGACAGCCCGCTCGCGGCTTCTGCGCCTTGGGTTCGGTCA  
5 AGTCGAATGTGGGTCAATTTGGATGCGGCAGCGGGCATCGCGGGTTTCATCAAGGCGGTGCTCT  
CGCTGTCCCATCGGACCCTGTTGCCAGCCTCCACGTCGACACGCCCAACCCGCAGATCCCGT  
TCGCCGACGGTCCGTTCCAGGTCAACACGAGACCCGGCCCTGGCCAGCTGCCGACCATCCCC  
GCCGCGCCGGCGTCAGCTCCTTCGGCATCGGCGGCACCAACGTGCACGCCGTCCTGGAAGAGG  
CGCCGAGTTGGCCGAGCAGCGGGGCGGCGCGAGCGGCAGCTGTTCTGCTCTCGGCGC  
20 GACTGCGAGCCGATCTGGAGCGACGACCGCGCGCTGGTCCGCCACCTGGCCGCGCATCCGG  
ACCTCGCACCAGATGACGTTGCCTTTACCTTGACGCGGGCCGCAAACCGATGACCCACCGTC  
GTTTCTGCTCGCCGCCGACCTCGCGGAAGCCCGCGCGTCTGGCCGAGCCCGATCCAGTCA  
AATCCGCCGCGGCGCGCGCGGACCGCTGCCAGGTCTGGATGTTGCCCGGTCTCGGCTCTCAAT  
25 ACCCCGGCATGTGTGGCGGCCTCTATCGCACCAGCCGGCCTTTCCGCGAGCAAGTCGACCGCT  
15 GTTTCGACCTCCTCGCGCCGCGTTCGGATTTGAAGCCCTCGCTCTTCCCGAGCCCGATCAGG  
CCATCGACGCATCAGCCCTCGCGGCCATCGACACCGCCAGATCGCCGTCTTCGTCTGCGAAT  
ACGCGCTCGCACGGATGCTGGAAGGCTGGGGGTGCGTCCGGATCGGCTGATCGGTTACAGTT  
30 TCGGCGAATACGTGGCCGCCTGCCTGGCCGGCGTCTTCTCCCTGCCCGACGCCTTGGCAATCG  
TCCGCGAGCGTGGCCGGATCCTGGCGGCGGCCGAGCCGGGCGCGATGGTCAGCGTGCCCCCTC  
20 CGGCCGAGCGCGTTCGCTCGCTGCTGGAGCCGCGCTTGCCCTGGCCATTGACAACGGCCCCCT  
CATGCGTGGTGTCCGGGCCGGTTCGAACCGGTGCGCACCTTCACCGCTCGCATGAAGCGGGACC  
35 GGGTCTGGGTGACGCCGCTCCAGGCCGAGCGCCCGATGCATTGCGCGCTGATGGCCGAGGCCG  
GCGGCTCACTGCGGCCATGTTGGCCGGGTTCGCCTGAATGCGCCGGAATCCCGATCTTAA  
GCAATGTTACAGGAACCTACCTAACCGACGAGCAGGCCCGAGACCCGATTACTGGGCCCCGTC  
40 ACCTGTGCGGCAACGTTGCTTCGCCGACGGTGTGCGAACCTTGTTGGCCGAGCGCGATCCGG  
25 TCTTCCTTGAATTCGGGCGGGCCGCGATCTGAGCTCCTTGGTGCGCCACCAGATGCCGGAAG  
GCGCCGACGAGCCGATCGCACTGATCCGTTCATCGGAAGATCCGGTGCGCGACGAAGACCTCC  
45 TGCTCGATGGCTTGGGCGGCTGCTTCTGCGTGGGGCGACCCTCCACGGGCAGGCCTTGTACG  
CCGGCCGAGGCTGCCGCCGCGTGCCGCTGCCCGGTTACCCGTTCCAGGGTCCACGCTGCATGC  
30 CGGCCCGCGCCGACTGCCCGCCTGGCGCGACCGACCGTGGGAGCGACCACCATCAGCTACC



5

72

10

15

20

25

30

35

40

45

50

55

GACCAGCCTGGAAGCGGGCGCCGCGCTTGGCGGCTGTCTGAATCGCTCGCGCCGCAATCCTGGT  
TGGTATTTCAGCGACGGCAGCGAATTGGCGGGCGAGCTGGTGGCCGGCCTGCGCGCTTCCGGTT  
GCCGACCAACCTTCGTCTGAAGGTGGGCTGGCGTTTCGCGCGCTTCGCGGGCGGCTTCCGCGCGA  
ATCCCCGCGAGGAACAAGATCTCGCACAGCTGTTTCGCGACCTGTTCGGCCGAAGCGATGCTGC  
5 CCACCCACATCCTGCACCTGCTCAGCCTGCCGTGCGCGGAGCGCGACTCGCCGCTGGCGCGCC  
TGGAGCACCTCACCGAGCTGGGCTTCCACCATCTGCTGGCCCTGGCCCCGCAACTGGAGGCGG  
TCGGCGCCCCGAGGTCCGCCTCGCCGTGGTGACAACCGGCCTGGCGGCGATTGGCGGCGAGT  
CCGAGCTGCGGCCCGAGGTCCGGCTGTTGCGGGGACCTGTCCGCGTGATTCCCTTTGAATTCC  
CGAATTGCGGCTGCGCCTGATCGACCTCGACTCGGCCGATCCCATCTGGCGTAGCGGTTGTG  
10 AGCCGTTGCTGCGCGAAATGGGCGCTGCCCCGGGACCTGAAGAAATCGCGCTGCGCGGCACCA  
GCCGTTGGGAGTTGGGCTACGAGCCGGTTCGAGGGGGCACCGTGAGCACCATCTCCTCGCGAC  
TGCGCGAGGGCGGCTCTATCTGATCACCGGTGGCCTCGGCGGCGCTGGGTCTGGCCTTGGCCC  
GTCACCTCGCCCCGAAGTACCGCGCCACCTGATCCTCGCTGGCCGCGGAGGCGCGCCGGCGC  
25 GCGAGCTCTGGCACCAGGCGCCAGCGGAGTTTCGTACCGGTTCGAGCTGCGATCGCACAGATGG  
AGGAGTGTGGCGCCCGCGTGATTCCCGTTCGCGCTCGACGTCACCGACGCGGACCAAGTGAACG  
CGTTGTTTCGCCACCATAGAAGCTACGGTCGGCAAGATTGAAGGCGTTTTCCACATGGCTGGCA  
TCGTTGACGGCGGCATCATTCTGAACGCGCACGCGCGCTGCCAGCGACGCCGTGCTGGCGCCCA  
AAACGGTCGGAACCTGGATTCTCGATCGGGCTCTCCGCGGCGCCGGTGGCCGCTTCTCTGGTGC  
TGTA CTCTCGATCAACGCGGTCTGTCGCGCCCTTCGGCCAGGTTGCCTACGCCGCGCCCAACG  
30 CCTTCTCGACGCCTTCGCCAGCGCCACGAACACGACGAGCGTCTTTTCCGCGTCAGCATCG  
GTTGGGACACCTGGCGCGAGGCCGGCATGGCCGTTCGATGCCGCCGCGCCGCGGCGACAGG  
CCCCGCTCGAAGGGCTTAGCGACGAGCAGGGCTTTCGCGCTGCTCGAAAGCGCCTTGGTTCGGTT  
GCGAACC GCGACTCCTCTGCTCTCCATCAGCGAATGCGCGCTCGACTAGCCGAGCATCATCGCA  
ACGGCGGCATTCCCCGGTTGCTCGGGCCCCGCGCCAACGAGGCGGGTGCAGCTGATTCCGGCG  
25 AGGAGGGCGCCACGCAAGACGCGTCGCCGGCCCGTCGCGCCCGTCCCGATCTGGTCTGGCCT  
TCGCGCCGGCCGGCAACGAGCTGGAGCGCCGGATCGTGGCCATCATCGGCGCCTACCTGCGGC  
TCGGTCAGGTGGGCGTCGACGACAACTTCAACGATTTGGGCGCCACCTCGCTCGACCTCATCC  
40 AGATCGCCCAACGCCTCGGTTCGCGAGTTGGGCCGCGATGTCCCTGTCTGCTCTACCAAC  
ACCGCACCGTACGCGGGCTGAGCCGCTTCTTCGGCGGGCGCGCTCCAATCCGCGGGTCCGGCG  
30 TCCCCACGGGCGCTGCCGCACCGGCGCGCCACGCGGGGGTTGCCACCCCGCCGCGGCCAC

5

73

10

15

20

25

30

35

40

45

50

55

AACCGTCGCGCCAGCACCTGGAAAAACGCCGTCAATTGAGGAAAAAGGGGGCCTTCCCATC  
ATGAGTGAAGTATCCATTGCCCCGGCTTGGACATCGCGGTTCATCGGCATGGCCTGCCGCTTT  
CCCGGTGCCCCGAACCTCGCCGAGTATTGGGCCAACCTGATCGAAGGCCCTCGAAACGCTCAGC  
TTCTTCAGCGAAGAGGAGCTGCGTGAGGCCGGCTGCGATCCGGTCCAACCTGGCCCAGCACAAAC  
5 TACGTGCGCACCAAGGGCCTGCTCCTGACGCAGACCGTTTCGACGCCGATTTTTTTGGTTAT  
TCCCCGCGGAAGCCCAGGTGATGGACCCCCAGATCCGCGTCTTCCACGAGGTCTGTTGGCAG  
GCGCTGGAGCACGCGGGCTACAACCCGCATCGCCACACCGGCACGATCGGCCTGTTCCGCCGGC  
GCCGCGCCCAACGTTTTTTGGGAGTTTCTCTCCTATCGGTCCGATGCCGCCAATTTAGGCAAC  
TTCACGCTGGGCCTGCACAACAACAGGACTACCTGAGCTCGCGCATCGCCTACAACCTCAAC  
10 CTGACAGGGCCCAGCTACACCCTGTTACCCGCTGCTCGACCTCGATGGTCGCCATCCACCAG  
GCCGTCCAGGCGCTGCTCAACGGCGAATGCGACCTGTGCATGGCCGGCTCGGTCTCCATTACG  
CTGCCACTGGTTGCCGGCTACACCTACACGCGGGCATGATCGTCTCGCCCGACGGCCATTGC  
CGCACCTTCGACGCAGGCGCCAATGGCACTGTCTACGGCGACGGGGCCGGCGTGGTCGTCTC  
AAGCGGGCCGAGGATGCGTTGGCCGACGGCGACCACATATTTGCGCTCATCAAGGGCTCGGCG  
15 CTCAACAACGATGGCAGTCGCAAGACCGGCTACACCGCGCCCAGCGTGCAGGGGCAGGTGGAG  
GTGATCCGCGCGGGCATGAACCTGSCGGAGGTGAGCCGGAGGCGATCAGCTACGTGAAACC  
CACGGGACGGGCACCACGGTGGGCGATCCGCTGGAGTTCGAGGCGCTAAAGGAGGCCTTCGGA  
GGTGGCTGCAAGGCCCTTCTGTGGAATTGGGTTCCGGTCAAGCCGAACATCGGCCATCTGGACGTG  
ACGTCGGGGATCGCGAGCTTCATCAAGCTGGTCTGCGCTGGAGCACCGCATCCTACCGCCC  
20 ACGCTCCAATTCCAACCTGCCAACCCGAAGATGGATGGTTCGATAGCCCCCTTCTACATCGTG  
GCTGAGCGCGAACCTGGCGGAAGATCTGCTGCCGCGTCGGGCCGGTGTGAGCGGTTCCGGT  
CTGGGTGGCACCAACGTCCACATGATTTTGGAGGAGTTTCAGCGGAACCGGCGGCGAACAGC  
GCGCGCACGCGCCACCTGACGGTGTGACGGCGCGGTGCGCGCAAGCCCTGGCGCAGCTGGCG  
GCCAACCTCGCCGAACACCTGCGCGAACACCCCGAGTTGGCGCTGGCCGATGTGGCCCATACG  
25 CTGCTGCACGGCCGCAAGCCACATCCATTGCGCGCATCCTGGTGGCGACCGATACGACGGCG  
GCGATCGACGCCCTTGATGAACGACCGCGATCCGCGAACGCGTTTCTTGAAGCGACCGGGCGC  
GGCGAGTCGGTGATCCTGTGTTTTACGAAACGCCGCGGAGCCGGAAGCGCCCGCTACCTC  
45 TGGGATCACGAGCCGCTTTATCGCGCGGCGGACGTCGTGCTTGGCTGGTGAGGTGCGCGAC  
CCGATCTGGAAGGCTGCTTTACTGCCCTGATCGCCGAGCAGGGCGCGGACGCCCTTTTGC  
30 CACCAATACGCGCTGGCCGGATGCTGCTGGCCATGGGGTTGACCCCGTCGGCGTTGATCGGC

5

74

10

15

20

25

30

35

40

45

50

55

GTGGGCCAGGGCGAGTGGGTAGCAGCGGCGCTCGCGGAGGTGTTCCCGCCATCGGCCTGCTTG  
CGCTGGATTAGGTTTCGGCGAACGGCTCCCGCAGCCGCGCGATCAACGGATTCCGTTTCTCTCC  
AATTTCTCTGGAAACTGGATCGTTGGGCGTGAGTTGGCCGACCCGGATTACCCAGAAAGCAG  
AAGGGTAAGCGCTGCATGAAGCGCCGTCGGTCCCAACCTCGGTTCAGCTGGTGCAGGATGGGGG  
5 CGATGGAACCGGCTCGGTTCAGCTCGTCGCGCGCTGCTCTTCGCGGGAAGCGGAGGCGGGACG  
GTGATCGCCCCGAGGGCGAGGTTTCATCTCGTCGTCGACGAGCCGGGCGCGGTGCGCGCCCAG  
TACCTGGGGGCGAGCTCGAGGTAGCGGTCCCGCGGCCAGTAGGGCATCGCGCGAATGACGTCG  
GCCAGGTAGGCCTCCGGGTTCGAGCCCGTGCAGCTTGACGCTCGCCACGAGCGAGAAGAGTTG  
GCCGCGGCGGAGGCGTGTCGTCGTCGCCAAGAAGAGCCAGGACTTTCTCGCAACCGCAATG  
10 GATCGCAGCGCTCGCTCGCTGGCGTTGTTCTCCAGGCGCAGCCGACCGTCGTCGAGGAAGCGC  
CGCAACGGCTGCTCTTGGTTGAGGGCGTAGCCGAGCGCGGTGGAGACCAGGCCGCGCTCGCGG  
GGACGAGCGTGCTCGGCCCTGGCCCCAGGCAAAGAACGCGTCGACCAGAGGGCGGACGACACA  
TCGCGACGCACCTTGCGCTGCGCGGGCGGCAGGTCCGCCAGCGCGCGATCGGCGGCAAAGAGG  
25 GCGTTGATGCGCCGACGCCCTCGACACCGAGCTCGTGCTTGACAGACCGCCGCTCCAGAAAG  
TTGGTACGGCAATGCGACCAGCATCCGACTTCGGTCGGGGGCGGACCGCGCTTCTCGTCGGCA  
GCAGCGCTCTTGGTGGTGTGCCGCGGAAGAGGGCGTCATAGATGGCGTGAGCGTCAGCTTGA  
ATATACCGAGAGAAGCCGCGGAACATCTCGCAGACCGCGCGCTGGTATGCTTGGGCTGGTAC  
TCGAAGAAGACGTGATCCTTGTCGCGAGGACGACGAAGAAGTGTCCTTGCGGCACGGCCCCG  
GGCTTCTTGTCCTTGCGCTCCTGGATGGGCCAGGCTGGACGGAGACCCCGGTGGCGTCCGTG  
20 GACAGGCAGAAGGCGGTCTCGAAGGCCTCTTTGCGCGCGGCCTCGACGATGGCGCCCAGGGTC  
GCACCGACGTCTTCGGCGTAGCGGCACATCGTGCCGCGATCGAGCGACGCGCCCTGAAGCTCC  
AGCTGCTGCTCCAGTCCGATAGAACGGGACGCCGAGCAGGTACTTGCTGGTGAGGATGTGCGCA  
ATCATCGACGGCGCGAGGAACGACCGCCGGAACAACCTCCTTCGGAAGCGGCGTCGTGATGAAG  
40 ACCGTGCAGGTCTCGCCCTTCGGCGCCGGCGGCGCGCTCGAGCGAGGGCGCGTCGAGCGCT  
25 GTGGAGGAAGCGCTCGGCTCGCCGGCCGCTGCCGTGTCTCCGGGCTGACGCTCGCAGCGGGC  
GTGGGCGTCGGGGCTTCTCTCGCAGCAGCTTGAGCGGGGCGCTTCTCCTCGCCCCAACTG  
CTCGCATCCGTGACGGACCGCTCGGCCTTGACACGACGCGTGCGAGCACGATCGGGCGCATT  
45 CCGCGCGCTCGTAGCCGAGTCGCGAGGTCTCCTCGACCCCGATGCGCGTCGCGCTCGCATCG  
AGCTCGGGGACGAGAGCTCGATGCGGACGACGGGCGAGTCCGACTCGGACAGGTGCGGACGG  
30 CCCTTGCCGCGGACCTTCGTTTCGGCCCCCTTGGGGTCGTGCTGCGCTCGCTCGCTGTA

5

75

10

15

20

25

30

35

40

45

50

55

TTGCGCTCGGCGGCGTCGAGTGCCTTCGCGAGGCGCTGGACCTCGAGGAACATCGAGTCGAAC  
GCCAGCTGCTCCGCGCTCACCTCGGCGCGCTCCGCCTTGGCCACGAACAGTCGACGTCGCAGA  
AGCTGCAGCTGCTCGAGCGCACGGGTGTAGGCGCGCCGAAGCTGCGCGAGCGCATCGCGCGCT  
CCCACGAGCTCGCTCTTTGCCGCGGCGAGCTCCGCTTCGAGCTGCGCGATGCGCTGCTGCTCG  
5 GCCGAGAGCGTCGGCTTGGCGGCGGCGTCGTGCACGACGCCGCTCTACGTAAGCCGCGCGTAC  
15 TTGTCGAGCGAATTCGTGCGGCTCAGTGSACGCGGCGCGGTGCGCGCCTTCGCGGTTTGGACG  
TGGGCGCGATCTCGATGCCGTCGAGCAGCGTCTCAGCGTGCGCTCGTCCACCTCGACGTGCG  
TGGCGCCCTCGGTCCGGGGGTCCGGAAGTGCGAACGCTCCGCGATCAAGGCGTTTGTAAAACA  
GGCAGATTCCACTGCCATCGAAGAAGAGAATCTTGATCGTGGTCCGCCGCTTGCCGACGAACG  
20 CGAACAGCGCTCCGCAGCGAGCCTCGTACCCACACGCTCACGGATGAGACCCGAAAGCCGCT  
10 CGAAGCCGTAGCGCATGTCCACCGGCTCCAGCGCAGCAACACCTGCACGCCCGCCGGAATCA  
TCGCCCCGCTCCGCCGAGGGCACGGACCACCTCCGCCAGCAGCGGGGTGCAACCCCGCGGC  
25 GACGCGCACCCGCGCGCCCGACCTCGACGACGAGCTCCGCAGCGCTGCTCGTCACGGCGGG  
CGCCTTCGGCACCAGGCGCAGAAAGCGCGGTGGCTCGCCCCGCGACAGCCGGCTCGACCAGCC  
15 GTGCAGCGTCGAGGCCGCAATCCGCGGCTCCGAGCGAACTCTCCGCCGTTTCACTACTCTC  
GCGCCACGCCCCGAACGCGCTCGGACCACATCACTTCGGTCGCCTTCGTCTTGTCTGTCACGC  
30 CATCATGAACTGGACAGCGCAGCCGGGTGAGACGGCGCTTCGCGCAGCGCTTACGCAGAAGG  
CGCGCCGCGCGCCATTGTCTGGATGCGGTGCGCGACTTCGCCCGCGATCGGCTGTTGCTGGAAC  
TGGGACAACCACTGGACGTAACGGCTGAAGCGAGCCAACGGCTCCAGCTCGCGCGGGGCGACC  
20 TGTTCCGGCGCCTACCAAGCGTTGGCCAGCTCTGGATCTGCGGCGCCCTGGCCGAACCGCCGC  
35 GACTGTATCCCGACGAACACCGCCGGCGCGTGCCGCTGCCGAGCTACCCCTTCGAGGGAAAGC  
GGTTCTGGATCGAGGGCTCGCCGTTTCGAAACCGCGCCCGCCCGCGCCTACCCCAACCCG  
CCGATTCGGGGGACATTCTCAAGGGCGACCCGGCGGACTGGTACTATCGGCCGCGTTTCGAAG  
40 CGGCGCCGCTCTTGCCCAGCCCGTTTCGAGAGCGAACC CGCGATTGGCTGGTGTTCGAAGATG  
25 AGCTGGGGCTCGGCGCCTGGCTGAGCGAGACCTTGCGCGACAAGGGCGCGGGTTCGCGACAG  
TCGTTTCGAGGCACCGAGTTCGACGCGCTGCGGTACAGCGCTTCAGCTTCGTCCCGATCGAC  
GGGACGATTACCGACCCCTGCTGCACGAGTTGAAGGCGCAGGGCATCGCGCCGTTCCACCTGT  
45 GCCACCTATGGAGCGTGACCGCCGCACCGGATGCCGAGCAGTTGCTCGACGTCAGCTTTCACA  
GCCTGGTCCATTGGCGGCCGCTTTGGGTTGCGTTGGCTACTTCCACGCCATGAAGTTGAACG  
50 TGGTCGCCCAACCGGCTATTTCGACCCCGAGTCGCCCCGAGCGCACCGAGCCCGCCAAGAGTCTGT

5

76

10

15

20

25

30

35

40

45

50

55

TGCTCGCGGTGACCAAAGTCCTGCCGCAAGAGGTGCCCAACGTTTGAACCCGCGCCATCAGCG  
TGGACCTGGATCGCTCGTTTCGACGCGGCGGCGCCCGCTGGGCGGCCAGTTTGTGGTTGAAT  
GCGGCGCGCCCGTCGAGGAAACGGTGGTGACCTACCATGGCGCAGCCCGATGGCTGCGCCGCT  
TCGATCGCGTTGCGGTGAATGGTCTCGGCCCGTTCCACCCCGATCAACCTGCGCCGCTGCTGC  
5 GCGAGCGCGGCGTGACCTGATCACCGGCGGCGCTGGGCGGCGTGGCTGGCCAGTTGSGCGGCT  
ACCTGGCGCGGGCGCTGCCGGGCGCGSTTGGTGCTCACCGCGCGCGGCCCGCTGCCGAGCGCG  
ACCACTGGGATCGGGAGTCGGCCGTGCTGTATGGGACGACAAGACGCGCCAGCGCATCGAGC  
TGGTGCGCGAGCTGGAGCGGCTGGGGGCGGAAGTATTGGTGGTGGCTGCCGATGTGCGCCGACG  
AAGCGGCCATGGCGCAGGCGATCGAAGCCCTCACTGGCGCGATTTCGACGCTTTGGACGGCTTGA  
10 TCCACGGCGCCGGGATCGTGCGGCTCGCGTCGGGCGCGCACGCCGATCGGGAGTATGACGCGGG  
CCATGTGCGAGGAGCAGCTCCGCCCCAAAGATGTTGGGCGCTCGACGTCGTCGACCGCCTCTGCG  
GCGATCGCGGTTGGACTTCGCGATTGCCATCTCGTCGCTCGCCCCGATTCTCGGCGGCGCTCG  
GCCACGTGCGCTACGCCCGCGCAACCTCTACATGGACGCGTTTCGCGACGCGCGCGCGCGCGG  
GCAACGCGCCTTGGATCGCGCTGAACCTGGCCGAGTGGGAATACGAGGGCCCGGCTACCTACG  
15 ACGAGCGGGTGGGCCGTTTCGCTCAAGCAGCTCGAGCTACCAACGAGGAGGGTATCCGCGTCT  
TCCAGACGGTGTGGCCTTGGCGCGCGCGGCGCGCTACAGCAGATCATTATTTCCACCGGCG  
ACCTCCAGGCCCCGCTCGACAAATGATTACATCAAATCCCTGCATCGCCGACCGGGGCGCGG  
TCCAGCTCAGTCGCCGACCGCGGCACCCAGGGCGGTTTCGGCTCGGAGCGCGCGCGCTTCG  
AGGCCGCGCTTCGCTGACGCGCTGGTGGGACTTCTTCGGGGTTGAAGAGGTGACCCGAACAAA  
20 ACTTCTTCGATCTGGGCGCCAGCTCGCTCGACTTCATCCACCTCGTCAGTCGCTTCAGCAAGG  
CCATCGAACAGCATGTACCGCTCGAGGCCCTGCTCGAACACTCCACCTGCACGACCTCGCCG  
CCCACCTCGCGGGCGACGCGAACACCGACGCCAGCGACGAAGCGCGCATTTCGCCAACGGCTGC  
AAGGCGCCAAGTCCGGCGACATCGCCATCATCGGCATGGCGGCGCGCTTCCGCTCGCGCCCCG  
ACCTGGACACCTATTGGCGCAACCTGGTCGGAGGCATCGACGCGGTCAGCTTCTTCAGCGCCG  
25 AGGAGTTGCGTGCTGCTGGCGTCAACCGCGGCCGAGATCCACCACACCAACTACGTGCCGGCCA  
AGGGGCGCTGCGCCGACAGGACTTGTTCGATGCGGCCTTCTTCGAATACACTGCCAGCGACG  
CCGAGCTGATGGACCGCAAAATCGCGTGTACACGAGGTGCTGTGGCACGCGCTGGAAGACG  
CCTGTTTCGACTTCAACGGCGATCACGGCCAGGTCGGCCTGTTTCGGGGCGCCTCGCCGAACC  
40 TGTTGGTGGCAGTTTCGTGGCCAGCTTTTCGAGGCCCGCAAGACGAGGGCATGTTACACCA  
30 CCCTGCTCAACGACAAGGACTCGATCGCGACCCAGATTTATACAAGCTCGGTCTAAAGGGCC

5

77

10

15

20

25

30

35

40

45

50

55

CCGCGGTACCTTGTTACCGGCTCTTCCACCTCGCTGGTAGCCGTTGACGCCGCTGCCGCT  
CGATCTGGTCCGGTCAATCGGACATCGCCGTGGCCGGCGCGGTCTCGCTGACTCTCCCCGATA  
AGGCCGGCTACATCTACGAAAAGGCTATGCTCTTCTCGGCCGACGGCCATTGCCGGGCTTTCCG  
ACGCCAACGCCACCGGCATGGTCTTCCGGCGACGGCGCCGGCGCGATCGTGCTCAAGCCGTTCG  
5 ACGGGGCCCTGCGCGACGGCGACCCGATCCATGCGGTGATCAAGGGCTGCGCCACCAACAACG  
ACGGCGACCGCAAAGCCGGCTACACGAGCGTCAGCGCCCAAGGCCAGGCCGAGGTGATCCGCT  
CGGCCCAGATCCTGGCCGACGTGGCCTCCCGAATCCATCAGCTACGTGGAAGCCCACGGTACCG  
GCACCAAGTTGGGCGACTCGATCGAGATCAAGGCGTTGAAGCAAGCCTTCGCCAGCGACAAGA  
ACGGATTTTGGCGCATCGGGTCGGTCAAGACCAACCTCGGTACCTGATGGCGGCGCGGGGA  
10 TGGCCGGCCTGATCAAGACGGTTCTCGCGATGAAGCACCGCCAATTGCCGCCATCGCTGCACT  
GCGACGAAGTGAACCCGACCTGGAATTTGGAGCGCAGTCCGTTCTACATCAACACCCGCTGC  
GCGACTGGGTGACCGGGCGGGCCCTGCGGGCCGGCGTGAGTTCTGTTCCGGATCGGCGGAA  
CCAACGCTCAGTCATCCTGGAGGAGCCCGGACGCGGAGAGCGGCACGCGCATGCCCACT  
25 GGAAATTATTGATGCTGTGCGCGGCCAGCGAGGCGGCGCTCGACCGCCAGGCCGATAACCTGG  
15 CCGACTACCTGGAGCGCCATCCCGAGCGCCACCTCAGCGACGTGGCCTATTCCCTCCAGACCG  
GCCGGCGCGTTCTGGCCTGGCGGCGCACGGTCTTATGCGAGTACCGCGAGGACGCGGTGACCA  
GTCTGCGCGAGCGACAGGCCAAGCTCGTCCAGACAAGTCCGCTCCGCTGGGACCACAAGGACG  
30 TGGTCTTCATGTTTCCCGGTACAGGCGCGCCAGTACCTCAACATGGGCCGCGACTTATACGTCA  
TGGAGCCGGTCTTCCGCGAGGTGATGACCGCTGCTTCGAGTTGCTGGCCCCCTTTGTGGTCCG  
20 AGCATCCGCGCCAGATCCTTTATCCGAGGGCGGGGTGTCGACCCTGCTCCACCGGACTGATT  
ACACCCAGCCGATCGTGTCTGCTTCGAGTACGCCCTCGCCATTGCTGCTCTCCTGGGGAT  
35 TGAAGCCGGCCGCGACCATCGGCTACAGCTTCGGCGAGTACGTTTCTGCCTGCCTCGCCGGCG  
TCTTCTCCCTGGAAGATGCGATCCGCTCGGTGACCGAGCGCGGTCCGGCTGATGGCGGCTTTGC  
40 CCGCGGGCGCCATGCTCAGCGTCCCTCGTTCCCGAATGCGAGCTGCTGCGGCTGCTGGACGGCT  
25 TCCACGCCCAATCGGCGGCCCATCTCGCGCTGGCCGTGACAAATGGCGCCTCCTGCATTGTGG  
CCGGCGAGCAGGCCGCCATCTCGGCTTCGAATCGATGCTTCGCAAGAAGCGTCTGTTGACCA  
TGCCGTCGCGGTGAGCACGCGCTTCATTGCGAGGTGATGACCGGCGCGACCGACGCCCTGC  
45 GCAGTATCCTGCGGAAGATCCCCCTCTCCGCGCCGACAATTCCTTCAATTCCTGCGTCACCG  
GCACCTGGATCACTGCACAGCAGGCTACGGATCGCGAGTATTGGGTGAACCACATGTGCGGGA  
30 CGGTCCGGTTCGCGGCGGGTCTGACCGAGCTGGGTCAAAACCGCGAGGCGGTGTTCTGGAAG

5

78

10

15

20

25

30

35

40

45

50

55

TAGGTCCGGGCCGCGACTTGACGTTGCTGGCCACCGCATCCTGGCCGACAGCGCGCCGTGT  
TCGAGCTGGTCAAGGCGCCCGACGGCGGCGACGACGATGGGTTCTCCTGCTGGATCGATTGG  
CCAAGCTCTGGAGGCTGGGGATTTGATTGACTGGGCCGGCTTCTACGCGGATGAGCGGCGGC  
GGAAACTCTCGCTGCCGGGATATCCGTTTCGAGCGGCGGCGCTTCTGGATCGAGGGCAACCCGC  
5 TGGAGATCGCCCGGCGAGGCCCAATGTCCAGGGGCGCTGGTCAAGGCGTCGGACATCGGCG  
CTTGGTTCTACGTGCCGCAATGGCGGCGGTGCGTGCTCGCCGAGCCGGGTACAACGGCGGCGG  
15 GC3CCGCCGTACGGCGGAGCAGGCACGCGTCGTGACCGAGCTACGGGCGGGATGCGCGTCCG  
CCGGCTTGGGACGCGGGGCTGCGGACTGAATGGCGGTGCCCCGTCCGAGCGTCCGAAAGAAA  
GTGTAGCGCCAGCCGGGTGACCA3CGCAGCGGCGCAGACCGGCGCGGACTGCCCGACACCGA  
20 CTGGGAGCCAGCGGCTGTGCCAAAGGACGGGCGGAGCCGCGGCGGACCTGGCTTATTTTCG  
10 CCGACGCCGCGGATTGGCCGAATCTTTCCGAAGCGGGTTCAGGCCCCGCGCGAGAAGCTTT  
ACCTGGTGGCTTCCGGCTCGCGCTTCGAGCGCTGGCCGAGACCCGCTTCCGCTCGATCCCG  
25 GGGCCAAGTCCGATCACCGCTGCTTTTCAAGGCGCTCGACGAGGCCGACATCCTGCCGACCC  
ACCTCCTCGACTTCGCTCGCTTGACTGCGCGGGGCCGACGCCGACCCCATGGACCAGGCCG  
15 GCTTCTTCGGGCTGTTGCACCTGGTCCAGGCGATGGCAGAGGCCGGCTACAGCCATCCCATTC  
GGCTGCTGATCGTCAGTTGCGGCGTCTACGATGTACCGGTGCCGAACCGCTGCAGCCGCGCG  
30 GGGCCACGATGATCGGACCGGCTCTGTGCATCCCGCAACAGTATCCGCACCTCGAAACGAGCC  
ATGTGGATTTGGGCGTGGTCCATGCCGACGAGCTCCACGCCGCGCGCCAGCTCGACAGCCTAC  
TTGCCGAATGCCTAAGTGCAACGGCCGAGCGCCAATTGGCGCTGCGCGGCCGACACCGCTGGC  
20 TGCTGGACTACGAGCCAGTCCGCTTGCCGCCGCTCGACCCGGGCGCTCTGCCCTGGCGCCAGC  
35 GCGGGTCTACTTGATCACCGGCGGTTTGGGCGGGATCGGCCGATCCTGGCCGAACACCTGG  
CCCGCACGACCTCGGCTCGCCTGGTCTTAATCGGCCGCGAAACCCTGCCGACCGCGACGACT  
GGGACGCTGGCTGAACCGCCCGCAACCGGTGACGCCACCCACGAACGGCTGCTGCACAAGA  
40 TCCGCGGATTTCGGATCTGGAAGCGCTAGGCGCCGAAGTCTGCTCCTCGCCGCCGACGTCG  
25 CCAACGAAGCCGCCATGCGCGAGGCCTACGATCGCGCCGAATCCCACTTCGGCACATCCACG  
GGGTGATTACGGCGCCGGCCTGATGGACGCGCAAAGCTTCTCACTGATCGACGCCCTCGACC  
ACGACCTCTGCGCCCGCAGTTTGAAGCAAAAATCCGCGGCGTCTGCGTGCTCGACCGGTTT  
45 TGGCCGACCGCACGCTCGACTTCTGTCTGCTGATGTCTTCCATCTCCACCGTGCTCGGCGGCC  
TGGGCTATTTGGTTACGCGCGGCCAACGCCTTCTCGACGCCTTCGCCAGGCGCGCAGCC  
30 GCGACGCCGCTTTCCCTGGCTTAGCGTGGCCTGGAGCCATTGGAAGTACTGGACCGAGCGCA

5

79

10

15

20

25

30

35

40

45

50

55

AGATGGACAACGAGGTCGGCGCCGTCATCGACAGCCTCTCGATGGAACCCGCCGAGGGCTTCG  
AAGCCGTCACCCGCGTCTTGGCTTGGGGCAAGGCGCCCCACATCGCCAACTCGCCCCGGTGACC  
TCGGTCGCCGCCGGGATCAATGGGTCAAACCTGGCCAGCCTGAAATCGGCGCACTCCAGCGAGC  
CCGAGCCGGCTAGGCATGGACGTCCGGCGCTCTCCAGCGAATGGGTGCGCGCCGCGCAACGTGG  
5 TCGAAGAGAAGCTGGTCGCCATTTTCGAGCAGGTGTTGCGCACTGCGGCACTGGGCATCGAGG  
15 ACAACTTCTTTGAGTTGCCCGGCGACTCGCTCAAGGCGGTGATGACCGCGGCCCGTATTCAAA  
AGGAGCTGAACGTGGAAGTGCCGCTGCCGACCTTCTTCCAGATGCCACGGTCGCTGGCCTGG  
CCGAGTTCTGTGACGCAAGCCAAGCGCAGCGGCCGGGAGACGATTGCGCGCACCGCGCCGCGCC  
CACATTACCCGCTCTCGECTGCCAGGGCCGCCATTACCTGCACTACCGCATGGACCCGCGTT  
20 GTACCGCATACAACGATCCCTTCGCCAACCTGATCGAGGGTCCGCTGGACGTGGATCGCGTGG  
AGCGCATCCTGCACACCCTCATCTACGCCACGACTGCTTCCGCACCTCGTTCCACTTCCGCG  
AGGGCGAGCCGGTCCAGGTGATTACGATCGGGTGGACTTCAACCTGGCGCGGATTACCTGCG  
CGCCCGAGGATTGCCCCGAACGGATGCGCGATTTCATCCGCTCCTTCGATCTGGAGCGACCGC  
25 CCGCCATGCGCGCCGGCCTCTTCGTACGGGGCCCGAGCGCCACGTGCTGCTAATCGATTTTC  
15 ACCACATTATACCGATGGCGTGTGCTTCGAGAACTTCGTGCGCGAGTTGCGGGCGCTCTACC  
GCGGCGAGATCCTGCCCCGAGCTGGAACCTCGAGTACAAGGATTTCGCGGTGTGGCAGCATGAGA  
30 ACCGGGGCCGCGCGCCAACAGCGACCAGGCCCGCTACTGGACCGAGCAGTTGGCCAATGCGC  
CCGGGCCGATCGAGCTAACCACCGATTTCCCCCGTCCAGTCGACGCGAGCTTCCGCGCGGACC  
GCGTGCGGACCGTGCTTGATGCGGAGCTCGTTGCTCGACTCAAAGAGCACGCGGCGCGCCTCG  
20 GCATCACCCCTCTATAGCTGTGCTGGGCGGATTCTCGTTATTGCAGCACAAAGCTCTCCGACT  
35 CGCACGACATCGTCATCGGTTGCCCCGTGCGGGGCCGACCCGGAGCGAACTCCAGGATCTGC  
TGGGCGCGTTTCGTCAACACCCCTGCCGATGCGCCACCGCATCGACCCGACCCATACCGCACGGG  
TCTTCTTGAGCAGGTCCACCAGACAACCTTGGCGGCCCTCAGCTACCAGGAGCACCCCTTTTG  
40 ACGAAATGGTGGCGACGCTCGGGTTGCGCGCGGATCCGGCTCGCAACCCGATCTTCGACACGA  
25 TGTTCTTGCTGCAGAACATGGCCATGGGTGCAACCACCATTTCCCGGTCTGCGGCTCTCGCCTC  
ACGACACTTTTCACCGCAAGGCATTGTGCGACCTGATGCTACAGGCGACCGAGTATGACTGCC  
ACCTGGAGCTGGTGCTCGAGTTGCCACCGACCTGTTCCGGCTGGAACCGCGCAAGTCTTGC  
45 TCGACCGCTACCGCCAAGTCTTGAGTGGCTGTTGGCGTACCCCCATGAATCGATAGACGATT  
TGACGCTCGCCGGCCACTTTCGCGAAGTCGAAGTGACGATGTCGGACGAGGGCGACTTTGATT  
30 TCTCAGATTTTGAACCCCGCAACGTGAGAAACCTATGGCGCGCCTGAGCCGCACAGATCTCCA



5

80

10

15

20

25

30

35

40

45

50

55

ACTCGCCATTCAACCAGCGCACCGTGGAGCGCGAATATTGGCGCGCTCTGTTCGAGCGCCATCC  
GCAACGGTCCAGTTTGGCGGGGGTCTCACCGCCCCGATCGGCGACGAGTCGACCCGCGAGAC  
CTTGTCATTCTGTCCTCGACGAAGATCCCCTTCGGCTGAGTAATCGTTCCCGCAACGCCTGCT  
CACGGTGTGGCGGCTGGCCTCGCGGCTTTCCTCCACCGCTGCGACGGCGCTGAGCGCTTAC  
5 CCTGGGGTTGGCCCTACCGCGCCAAGCCGATGACCATCACCCGATCCTCAACAGCTTGATCGC  
15 GCTGGGGTTCGCGGTGACTCGAGTACGACCTTCCGCGATCTGCTCTATGCGCTTCGATCCGA  
ATACCACGAGGCGATGCGCCACGCCAACTTTCCGCTGGCGACCTGGTGGCGGGCTTACCCGG  
CGGAACGGCGCGCTTCGACGTGCGCTCAGCCTGGACCCCTTCACAGACGGCGATTCTGCTGGA  
AGACCACGCGATCGGCGCTTGTTCGGTTTCGATTGAGGGTGAGCGCTCACCTGCCGATT  
20 GCGATTGACCCCTGCGCGCTATGACCTGCCGCGATCGAAAACCTCGCCGATCGTTTCCGCCG  
CTTCCTCACGCGCTGTGCGGGACGCCTCCACCGTCATCCAGGCGCTGGACCTTTCGCTGCC  
AAGCGATGAATCGGTGTGGCGGTCACTGAAGGCGTGGCGCGGGCTATTGCAAGACCTGAC  
25 GCTAGACCGCGCTTCCGCCGAGGCCGCGCAAACGCCGATCAGCCGGCGATCACGTTGAA  
CGGGGACGTCCAGAGCTACGCCGAGGTGACCGCCGAGCGACGCGCTGGCCCGCCACCTCCG  
15 TCGCCACGGCGTTCGGTCCGGAACGATTGTGGCCGTCAACGCCCGCGCGGGCCTAATCAGCT  
GACGGCCCTGCTCGCGGTCCATAAGSCCGGCGCGCTACCTGCCGATCGATGCCGAGGAGCC  
30 GGCTGCCCGCCAGCAATTCAAGGTGCGCGACAGCGGGCGCGGTGGCACTGGAGCCGTGCC  
GGACCAGGCGCTGACCGTCACCGACCTGCCGCGGTCTTCTTGACGATGCCTCGCTCTTCGC  
TGACGGCGGGCTCGATGTGCCGCGCGCGCGGACTCGCTCAATCCGGCCTATGTGATGTACAC  
20 GTCCGGCTCGACCGGACAGCCCAAGSGTGTGGTGGTTCCCCACCGCGGCGTGGTCAATCGTTT  
35 GAATTGGGGGCGAGTCCCGTTTCCCGCTGGACGAACCGGACCGAATCCTCCAAAAGACGCGCT  
GCTGTTGACGTGTGGTCTACGAGCTGTTCTGGGGCGCATGGAGCGGGGCCACCTGACAT  
CCTCGAGCCCGGCGCGGAGCGGACCCCGACGAGTGCCAGGGCCCTGGCCGAGCGCGCCAT  
40 TACCGTATGCCATTTCTGTGCTTCGATGCTGCTCGTCTACTTGAAGTCATGCGGCGGCACCA  
25 TCGCGCGCCCGTGCCGGACCGGCTCCGTTACGTCTTCGTCACTGGCGAGGCCCTCGAACC  
CCACCTCGCCGGGCTCCAGCAGATTGGTTCGGCGCTCGGCCGACGATTCCCTCGTTAATCT  
GTATGGACCAACCGAGGCCTCGATCGAAGTCTCTGCTTCGCTGTCCCGCCGACCATGTGCC  
45 GCGCCGATCCCCATCGGGCAGCCGATCGACAACGTGCACTGCACGTTCTCGACCGGCGCGG  
CCGTGCGCAGCGCCCTATCTTCTGCGAGCTGTTCTGGCCGGCGACTGCCTGGCGCGCGG  
30 CTACCTCAACCGTCCCGACCTGACCGCGCTCCACTTCGTGCCAATCCCTTCGGCAACGGCGA

5

81

10

15

20

25

30

35

40

45

50

55

GCGCATGTACCACAGCGGCGACTTGGCGCTCGTGC GCGGCGACGGCCAAGTGGCGTTTCTCGG  
CCGCCGTGACCACCAAATCAAAATCCGTGGTCAACGGGTCGAACTGGGCGAAATCGAGAGTCA  
TTTGCGCGGGCTCGAAGGCATCGCCGCCGCCGTCGTCCAGGCCGAGTCGCAGCACCATGAAAC  
CCTGCTGCACGCCTACGTCGTACCAACGACGCGGGCCTCAATGCGGCCCGGCTGCGCGCCG  
5 CCTCGCTCAACATCTGCCCAGTACATGATTCCCCAGCGCTTCTCGCGGCTGGCCGAGTTGCC  
GCTGCTGGCGGCAGGCAAGATCGACCGCGCCGCCCTCGCGCAACGTGCAACGCCGCTCGCCAG  
CGCGCGCCCTTTCGTGGAACCCAGCGGGCCCAACCAGCAGCGTATCGCAGAACTGTGGCGCCA  
GGTCTTAGCGGTGCGCGAAGTCGGCGCCGAGGATCCCTTCTTCAGCATCGCGGCAACTCGCT  
CAATGTGCTCAAGCTCAGCGCCGCGCTGAGCGACGCCTTCGCGCGTGACATTCCCATGCCGGC  
10 CCTGTTCCAATACGACACCATCGCCGCCAGGCCTCCTGGCTCGACGGGCAGGTTGACGAACG  
GGCCCAATCCGCCGCGCTCGACCGGCGAGGCCGCGAGGCGGCGCTGACCCCTCAAGAGACCGT  
GGCCATTTTGTAGGGATTTCATGACGAACCATGACCATCACGAGGAGAGCAGCGGCCTGGAGA  
TCGCCGTCATCAGCATGGCCTGCCGATTCCCGGGTGCTGCCGATTGCGACGCATTCTGGGAAA  
ACCTGATCAACGGGACCTCCTCGATCACCATTTCAGCGACGACGAGCTGATCGCGGCCGGCG  
15 TTGACGCGCGGACCTGACGCGCAGTACGTGCGCGCGGCCGCGCCAGATCGATGACGCCGAAC  
GTTTCGACGCGGCCTTCTTTGGGTACTCCAGCGTGAGGCCGAGCTGATGGACCCCCAGTTCC  
GCCTGCTCCATGAATGCGCCTGGTCTGTCTGGAACAGGCCGGCATCGATCCGCGCGTCGAAG  
CCGCGCCGATCGGGCTGTATGCCGCGCAGCCGACACACCTACTGGAACGCGCTCTCGTCGC  
TCGACCGGGGCTCGGCCGAATCGGAGCAATTGCGCGCCGAACAACCTTGCAACCGCGATTTTC  
20 TGTGCACGCTGGTCGCCGCGCGCTCAACCTGAAAGGCCCGCGGTGGTGGTTCAAAGCGCCT  
GTTTCGACCTCGCTGTGGCGGTCCACTCGGCCTGTCTGTCGCTCCTGACCGGCGAATGCCGAG  
TGGCCTTGGCCGGTGGGGTGGCGCTGCGCTTCCCACGCCCGAGCGTTATCGCTACGAACCTG  
GCATGATCTTCTCGCCCGACGGGGTGTCGCGGCCGTTTCGACGCGGGCGCTAACGGGACGGTGC  
40 CCGGCGAAGGCGCGGGGCTGGTAGCGTTGAAGACGCTGAAACGTGCCCTCCAGGACGGCGACA  
25 CGATCCACGCCGTGATTGCGCGACCGCGGCAACAACGATGGTGCCCGCAAGACCGGGTTCA  
CCGCGCCCGACGCCACGGCCAAGCCGAAGTCATTGCGACGGCGCTGCGCCTGGCCCGGGTGC  
CGGCCGAATCGATCGACTACGTGAGGCCACGGAACCGGCACGCCGCTAGGCGACCCGATCG  
AGGTAGCCGGCTTGGTGGAGGCCCTTCGCCAGCGAGAAGCGCGGCTATTGCCGGCTGGGCTCGG  
45 TCAAATCCAACCTTGGTCATCTGGACACTGCTGCCGGCATCGCCGGCCTGATCAAGACCGTGC  
30 TGGCGCTCGAGCACGCGCACATCCCAAGTCTGCCACGTGCCACGCCCAACCCCGCGCGC

5

82

10

15

20

25

30

35

40

45

50

55

GCCTACACAAGACGCCTTTCCGCATTGCCGCCGACGGGATGGCCTGGCCGCGGCGTATGGCGA  
CGCCGCGGCGGGCGGCGGTGAGTTCTGTTCCGCATCGGCGGCACCAACGTCCACGCGATTTTGG  
AGGAGGCGCCGCCCGCGCGCCCGAGCTGGCGGACGGGCGCAGTCAGGTGTTCTGTTCTCTCCG  
CCAAGGACGAGGCGGCGCTGGACCGTGCCCTTGCCAACTATGGTGCGGCCTTGEAGAAGCGCG  
5 GCGACCTCGCGGCGGGCGGCTGGCCTGGACGCTCCAAAACGGCCGGGCCGATTTCGAATGGC  
15 GAGCCAGCGCGGTGGCATCCGACCTCGACGAATTGGCGGGCGCATTGCGCGGCGAGCGGCCCCG  
GCGCCGTCAAGAAAAACCGAATGGCGCGCGAGGATAAGCCGGTGGCGTTCTTATGTTCCGGGGC  
AGGGGAGCCAGTACCGTGGCATGGGCCACGACCTGTACCGCGAAGAGCCGCGTTTCCGGCACC  
ACCTCGACGCCTGCCTCGCCATCCTCGCCGAACACAAGCCCGAGATCGACTGGCTGGCGTTGC  
20 10 TGGGCTACCGCGACGAGGACGAGCCAACCGACAGATCGGGACGTCCTCGCAGGGCCCCGAGCC  
GGTCAGCCGCATCGAACCCAGCGGAGCTCCTCGACAGCACCGAATTGCCCCAACCTTTGCTTT  
TCTCCATGTCCTACGCGCTCGGTGCGCTGTGGCTCGACTGGGGCGTCCGACCCACGGCGATGA  
25 TCGGGCACAGCCTGGGCGAGTACAGTGTGTCATGTATTGCAGATTTCTATGCACTCGATCAGG  
TGCTGCCCTTCATTCTGACCCGCGGTGAGTCATGGCGCAATTGCGGCGCGGCTCGATGTTGG  
15 CCGTCAGCGGTGACAGCGTTCTGATGCGCGAGCTGATCGCCGATGCGCTCGATTTGGCGGCGA  
TCAACGGCGCTGACCAATTTGTCTGGAGCGGGCCGAGCGAGGCTGTCCAAGCCGCGGGGGTCC  
30 GACTGCGCGGCGCCGGCCTGCGTGCCACCGAGCTGAACACCTCACACGCGTTCCATTAGCCA  
TGATGGATCCCATTCTGGAGGAGCTAACGGTTGCCGGTTGCGGACTTCAGGTGCGTGTGCGGA  
CGATTCCGGTCTGTTTCATGCGTTACCGGAACCTGGTTGACGGCGAAGCAGCTGGCCGATCCGC  
20 GCTACCACGCGCGTCACGCGCGGAACCGGTGCGGTTGCGGCGGGCCTAGCGACGCTGACAG  
35 GGGAGGAGCCGCCGCTGATGCTCGAAGTGGGGCCGGGCTCGACCCTGGCGGCTTTGGCCCCGG  
AGCATTGCAATGCCCCGCTCCCGGTGCTCACCAGCCTGCGCCACGCTCGCCAGGCGACGCCCCG  
ATCGCCAATACCTGCTCGAAACGCTCGGCTGCCCTTTGGCGACACGGGGTTTCCGTCGATTGGG  
40 GGGCCCATGCCGACGTTTCGCGACGCTTGTTTCGCTGCCCGGCTATCCCTTTCCGGCGCGG  
25 TCGCCCGCTTAGCCGGCGACCCCTCCGCTGCTGGCCGGAGCCCGCGCGTCCCGCCCCCGT  
CGGGAACGCGCCAACCTCAGCGCGGACGCGCGGACCTCCCGAACACTCCGGAGCCGACATCCG  
GCGCCGTGTCGGCGATCAAAGCGCCAATCGCCGCCGCGATCCCGGCTCTATCGCCTCTCCT  
45 GGCGCCAGGCGGGAACGGCGCCGCTCGGTCCGCCCCGATCTCGGTCCGCCCCGCGACTGGATCG  
TCTTCGCCTCTGATTCTCACCTGCTCCAGGCGCTCAGGGCCAATCTCGGGACGCGCGCTCAGC  
30 GGGTGACGCTGGTGACGCCGGGCCAGGAGTACGCAGCCGAGCCGTCCGGGTTTCGGCTGCGGC

5

83

10

5

15

20

10

25

15

30

35

40

45

50

55

CGSACCAGATCGACGATTACCGCGCCCTGTGGGCGGACTTGGCGCAAACCGGTATTGTGCCAC  
GATACATCGCGTTTCCTCGCCCCGTTTCATGTACCGGGCGCGCATGGCGGGCGATGCCTCGACCC  
TGGACGAAGTGCGCGAGGGCGGCTTCCTGCCCCGTGACCCGCTTGATCCAGACTCGCCCCGCCAG  
GCGGACCGAGCGGACTTCTAAGCCTCACGATCGTCACCCCGCGCCCTGGCGCTGGGCGACG  
AAGCGACGCGCCCGGAATGGGCAATCCTGCACGGGATGGTCGCCGGCTTAAGCCGCGATTATC  
CCGAATGGCGCTTCGTCTCGATCGACGGCGGCGACCCATCCCCGCATCGGTGCGAAGGTCTGG  
CCCGCTTGATCGCGCTTCATGCGGTTCGACGAGGCTGGCCCGACCCGCTTGGCGCTGCGCGGCC  
TTCACGCTTGGGTTCCACAGTGCGAGCAGCTTCAGCCGGCCACCATCCCTGGGGCGGGTATGT  
GGCGCGAGGGTGGTGTGTACATGATAACGGGCGGATTTCGGCGGGATCGGTCTGGCGCTGGCCC  
GCGCCCTGGCTCGAGAAGCTCGCGCCAAGCTGATCCTGGTCGGCCGAAACCTGCCACCGCGC  
CGATCGATCTCGAGGCTTGGGACGCGCCCGCTTGATTCTCACC CGCGACGTGCGCGACGAAG  
AGGCCATGCGCCGCTCTTCGATGCGCGCACGCCGCTTCGGCGCCATCGACGGCATTCTTC  
ACCGCGCCGGTGTCCCCGCTGGCAGCCTGTTCGCCAACCAATCGGACGCGGCCTTCGAAGACG  
TGCTGCACGCCAAGGTTTCGCGGTACCCCTCGTGCTGCAAGGCCTGAGGGCAATCGATGCGCCGC  
TGTTGCTGATGTCTCGCTGGACGCCTGGCTTCCCGGTCCCGGTCAGACCGCTATGCGCGCG  
CCAACGCCTTCTCGACGCCTTCGCCAGTCTGCGCCGGCGAGAGGGAGAGCCGGTGTACAGCG  
TTGGCTGGGACAGTTGGTGCAGGTGGGCATGGCTGCTCGGGTCGCTGCCCCGATCGGCCGACG  
AACGCGGCCGCTGGCGCGCGAGGGGATCAGCCCTCGCCAGGTTGGCAGGCTTTGAGCCGGG  
CGCTCGCCCTCGACCCCCCCCACCTGATGATCTCGCGACCGACCTGACCTCGCGCTGGCACA  
GTGATCCAGCCCTACGCCGGTCGCTTCGAGCGAACCAGAGGTGGCGCTGCCGCGCTGGACCG  
CATCCGCTGCCAAGCCGTTCATCGACGCTGTTGGTGCAGCACTTCGCCACCGCCGCGTGC  
CTCCCGATGGCAACTTTTTTCGAGCTCGGCGCCAGTTTCCTTCGACATCGTCCAGCTCAGCGCTC  
GACTTCAACAACAGTTTCGGCCGAGATGTGAGCCACACCGTGCTCTACAGTCATCCACCGTCG  
CCTTGCTGGCCGGCTACTTCGCCAATGACCCGACGCCGTCCGGTGCTGCTGCCGACGAACGCG  
ACGAAGCGGTGCGTCGCGGCCGCGACCTCTTGAAGAGCCGCGGCGAGGAGTATGACCGTGGA  
GCACGAAACCGGATTTCGAAATCGCCGTTCATCGGGCTGGCTTGCCGCGTTCCCGGCGCTGCCGA  
CGTGCCGCTTCTGGCGCAACCTGGTCGAGGCCAAGGAGAGCGTGCGCTTCTTCGAGGACCA  
CGAGCTGCGGGCCGCGCGGTGCCCGAGGAGATCTTGCGCTGCCCAACTACGTGAAGGCCAA  
GCCACTGCTCGCTGATGGCGAAGCTTTCGACGCGGACTTCTTCGGGTTCCATCCGCGCGAGGC  
CGCCTACCTGGACCCGCAAGTTTCGCTCTCGACGAATGTTTGTGGACCGCGCTGGAGGATGC

5

84

10

15

20

25

30

35

40

45

50

55

CGGCTACGATCCCGCGCAGTACGCCFACCCGATCGGGTTGTTTCGCGGGCGTCTCCAGCAATCT  
CTCGTTCCTGTTTCGACCGCATCGATCCGCGCGACTCCCCCTGCAGAAGCGCTATGTGGCCGA  
GCTGAACGCGGCCTCCTTCGCCACCCAGATCGCCTACCGGCTCGATCTGAAGGGGCCGCCAT  
TTCGATTCAAACCGCCTGTTTCGACGTCCTGGTGGCGATTACCTGGCGGCGCAAAGCCTGAT  
5 CGGCGGCGAGTGCCACATGGCCTTGCGCGGCGAGCGACCTTGGAGGTCCCCAAAAAGCCCGG  
CTATCTCTACCGGAAGGCTACATCAACTCGCCGGACGGCCACTGCCGGGCCTTCGACGCCGA  
CGCGGCCGGCACCATCTTCGGCGACGCGTTCGGCATCGTCTGCTCAAACGCTACCGCGACGC  
CCTACGCGACGGCGATCACGTGTACGAGTGATCAAAGGCTCGGCGATCAACAGTGACGGCCA  
TCGCAAGGTGTCTACACGGCGCCGCGCAAGAGCGGTCAAGTGGCGGTGATCCGCGCTGCGCT  
10 GGCGGCGGCCCCAGGTAGAGCCGCAACCATTCGCTTCGTTCGAGGCCACGGGACCGGCACACT  
CGCCCGCGATCCGATCGAGGTAGAGCGCTTGACGGAGGTCTTTGCCGAAGCGGTTCGCGGTAC  
CTGCGCCCTGGGTTCGGTGAAGACCAACATCGGCCACTTGGATGTGGCGGCGGCGTGGCCGG  
TTTCATCAAGGCGGTCTTGCGCTTCGAGCGGCGCTCCTCCCGCCAGCCTTCACTTCGTCCG  
25 GCCCAACCCGGCCATCGATTTCAACGGGCCCTTCTACGTTTGTTCGCCAAATCGAGCGGTTGAC  
GGAGAACGGGCGGTTGCGGGCCGGGTGAGTTCCTTTGGCATTGGCGGCACCAATGCCACGT  
GATTCTGGAGGAAGCGCCGGCGCCGAGGCGAGACTGCGGCGCGGAGCCCGCAGGCGCGAG  
30 TCCGTTCTGTTCCTGCTATCGGCCAAGACGCCGATGCGCTGGCAGGCCGTTGCCACGACCT  
TGCCGACCACCTGCGGGCGCACCCGAGCTCCTCCTGGCCGATGTGGCCCTCACTCTGCAGAT  
GGGGCGGGCGTCTGTCGCTACCGCCATGTGGTCCAGGCTGCGACGGCGGAGGAGCTGATTCCG  
20 CGGTCTGGGAGCGTTCCGACAGGAGTCCATCCGCAAGAGGCGGAATCGAGTACAATGGGTGTT  
GGCAGGCGAGGCGATGTGCTTGACGCGGTTTGGCGCTGTACGCCGATTGGCCGGTCTATCG  
GGAGCGGTCGACGTCTGTCTGGCGATCGTCCGCAAGCTGCGCCAAATCGACGGCCGGTCATT  
CCTACATGAGTGGATCGAGCGACCGCGCGAGGTTCTGCGCAATGGTTCGACGGCGCTGGCGTT  
40 CATGTTCCACTGCGCGCTGGCGCAAGCCCTGAGCCAGGCCGCTGCACCCGACGCGCATGTG  
25 GAGCCGTGGGCTGGGCGACAGGTCGGCGTGGTTTTGGCCGAATCCCTGTGTTGGAACAAGC  
GCTGGCGCTGGTGTGTGCCAGACACCGGTTCCCGGCGATGCCACACCTCAGCGCAACGCTT  
GGTTCGACACTGGAAGGCTGCCGTTTTCTGTCCACCACGATTTTTGATTTCCGGCAGACAGCTC  
45 GGGTCGACCCCTGGACCTCGCCGAATTCGCTCATGTGATTTTTGGTGGGTGGCCAAAGCGC  
CTCGCCCAATGAGGCGGAGCTGCGCTCATGGAGCGACGCCGCGCCGAGCTGGTGACCTTGGC  
30 GATCGGCCCATCTTTCTCGAGGCCGCTCCGGGACGGTGGGTCTGGCGATCGACCCCAAGCG

5

85

10

15

20

25

30

35

40

45

50

55

ACCGATGACCTGTGTTTCAGCGCACGGTGGCCGCGTTGTGGGAATGGGATGTGACGTGCGCTG  
GGCTGCGTTCACCTCGTCGACCGGGCGTCGGGTTCCCCTGCCTACCTATCCCTTCGTGCGGGT  
AATTCACGATCGGCGACCCCTTCGCGGAGCAGGCGCGGAGGATGACTTGATTGCGGCGAG  
CGCTTCGCGCTCGGCCGGATCGCCGCCCCGAGCCGTCGGCAAACCTCGGCAGCGGAACGCCACG  
5 CGCCAGTCAAGCATCGCTCGGCAACCACACCGGCTCCGTCTCATACGTGGCCAGCGTGGC  
CGTGGCCACCATTTCTCGAAACCGTCCGTGCCTATTTTCGGGTTTCGCCGCGTGGTTCACCGA  
CGCCTTCTTCGAATTGGGCGCGTCTCGCTGGATTGTTGTTCAACCTGGGCCAGCTCCTTTCCGA  
TCGTCTCGGCCGCGAGGTTCCGACCCTGCTCTCTACGACCACCAACACCGGACCGTTGGC  
GCTGSCCTGACATCCGCGGCGCTCAGCGCAGAGGCGCGCCCTTAAGGGGCGGTTCATCGCGC  
10 ATCGACTTCGGGCACAGCGCGAGCTCGGCCGCTCCACCGACCGACGTTCCCGGGGACGC  
TCACTCGCAGCCAGCTTCGTTTCGCGAGCAGGACATCGCCATCATCGGGATGGCCTTCGGGG  
ACCGGGCGCCGACGACCTGGACGCGTTCGGAACAACCTGGTGAAGGGTTCGAGTCGATCAC  
CTTCTTCAGCGAGGACGAGCTGCTGGCGGCGGGCTCCCCGCGAACATCTGGCCTCGACGCG  
25 CTACGTGCGGGCCAAGGGGGAAGTACTGGGATGATGGATTTCGAACCGGAATTTTCGGTTA  
15 TTCGGCGCGGAGGCGGCGGTTCATGGACCCGAGTTCGCGTGTTCACGAATGCTCCTGGCA  
CGCACTGGAGCACGGCGGCTACGATCCGACCCGATGCGCGGCATCGATTGGCGTCTACGCCGG  
CGTGACCAACCACCTGCCTTGGCTGATGCGAACTTTGCCGCACCTGACCGAGGAGGCAATT  
30 CGGCGCGCTGCTCCTCACCGACCGGAGTTCCTCGCACCGCTGCTCTCCTACAAGGTCGGCCT  
GCGCGGACCCGCTATTTTCGCTGCAAACCGCCTGTTTCGACGTCGTTGGTGGCGATCGGCACGGC  
20 CTGTCGCGAATTGCGCGCGGCTGCTGTCAGATGGCCCTAGCGGGCGGCGTGACGGCCAGCAT  
35 CGA3CGCTGCGGCTACTTCACCAAGAAGGCTACATCCTCTCGCCTGACGGCCACACGCGCAG  
CTTCGACGCGGCGGCGCGGCGGCGGCGGCTTCGCGGACGAGTCGGCATGGTGTGCTGAAGCC  
GCTGGCCCCAAGCCTTGGCCGACGGCGACACGATCCACGCGGTGATCAAGGGAATCGGCATCAA  
40 CAACGACGGCGCGCGCAAGGTCGGCTTCACCGCACCTAGCCGGGCGGTCAGACCGAGGCGAT  
25 TCGGGCCGCGCTGCGCGACCGGGGTGGCGTCGAACCGCGTCAGCTACGTGGAGGCGCATGG  
AACC3CGACCAAGTGGGCGACCCGATCGAGGTCGAGGCTTGACCCAAGCCTTTCGCGCCGA  
AGCC3ACGGTCCGCTTCGCCCCGCTCCTGCCTACTCGGCTCGGTGAAGTCCAACGTGGGCCA  
45 CCTGAACGCGCGGCGGCGGCTGGTCTGGTAAAAACCGTCTGGCGCTCCAACACCGCCG  
CCTGCGGACCGCCTGTTCTACCAAGTCGCGCAATCCACACATCGACTTTCGCGCGAGTCCGTT  
30 CC3GTTGAACGGCCAGACTTCGATTGGGTCGCGCCAGAGGGGACGCGGTTGCTGGCGGAGT

5

86

10

15

20

25

30

35

40

45

50

55

GAGTTCGTTCCGGTATCGGGGGAACCAACGCCCACCTGATCGTCGAGGAGGCGCCGAAAGCGCT  
ACCGACGACAGCGGCACCTCTGTGACGGAGCCGAATGACCTCGACGCGGGCGACGCCGACGG  
CCTAGTGCTGCCGATCTCGGCCCGCACGCCGACCGCCCTGGCGCACATCGCGACCAACCTCGC  
CAATCACCTGGAACGACATCCGACCATCGCCCTGGCCGACGTGCGCCCTGACCCTTCAGCTGGG  
5 CCGTCGCCAATGGCCCCATCGCCACAGCCTGATCTGCCGGAATCGAACGAGGCGATCAAGCT  
GCTGCGCGCCGTCTGCTCACTCCGCGGAGGTGCCGCCAGCTCAGGCGCCGGTCTCGGATGCGCC  
GCGCTGTGTTTTCTTTTCCCGGCCAGGGCGCCCAATACCCGAGCATGGCCCGCGACCTGGT  
TCGAAACTGTCCCGACTTCGCCCTGCACCTGGACCCCTGCCTCGACCAGTTGGCCGAACCTGCT  
TCCCGAAGATCCGCGTTGCATCTGTTCGGCGATGGCCCCGCCGATCGGCTCGACCAGACGGC  
20 CTACACTCAGCCGCTGCTCTTCTCCGTGTCCTACGCCTTGGCGCGCTGGTTGGGCGATTTCGG  
CATTGCCCCGATGCGATGATCGGCCACAGCCTGGGCGAATACGTGGCGGCCTGCTTGCCCGG  
GCTTTTCTCGCTGAGCGATGCCCTGCTGCTGGTGAGTGAACGCGGCCCGCTGATGGGCTCGGC  
CGCGCGCGGAGCGATGCTGGCCGTCCCTTGCCCGAATGGGAACCTGGAGGAACGCCTGGAGCT  
25 TCTGCCGACGACCGAATCAGCATCGCGGCGGTCAACACCGCCGAGAGCTGCGTCATCGCGGG  
15 ACCCAGCGAGGCGATCGAGCGCTGCGCCCAGCGCTGGGCCGCGCAAGGCCTGACCTGTACGCC  
GCTGCGCAGTCCCACGCCTTCCACTCCGCGATGATGGAGCCGATTGTGCAACCTTCGGCCA  
TGCTTGGCACGGGTCACTTTCGCGCCGCCGCGCGCGCTGGATCTCGAACCTCGACGGCAA  
GCCGATCGATTCCGCGGCGGTGATGCAGCCCGACTATTGGGTGCGCCACCTGCGCCAACCGGT  
CCGCTTTCACGAGGGAAGTCACTGCTTGGCCGAGGACACCCATGCTTGGGTGCAAGTGGG  
20 TCCCGGCCGAACCCTGTCTCTTCTGTCGCGCCGCCACCCGGCCTACCGTCACCAGCCAATCGT  
CAACCCCATGCGCCATGCAGTCGAGTCGACGGGCGACGTGCGCCGGTGGCGCCAAGCGCTGGG  
CGAACTATGGCGGGCCGGCATGCCGGTGCCTGGGAGCGGCAGCGGCGCGGCCGCGCATGCCGG  
ACGACGTGTGCCGCTGCCGGGTACCCCTTCGAGCGGCGGCCCTTCGCGGCCCGAAGACCGGT  
GGAGCTGGCGCAGCCCGCGCCCAAGGCGGAGCTGGTGAAAAACCCCGATCCCGCGCGGTGGCT  
25 GTACCGCCGCTCTGGCGCCCTGCCAGGCTGCGGCCGGCGGACTGGCGGTGCAGGCGACCGT  
TCTGGTCTTCGGCGACGGGTCCGAGCTGTGCCGCGCGCGGTGCTCAGGTGCAGCGCCAGGG  
GCTGAAGTGCGTCTCGATCACCGGGGCCGCAATTTCGCGGGGAGAGCGACATGCGCTTCAC  
40 GCTTGACCCCGCTGATCCGCGCCAGCTCGACCAGCTCTTCGCGGCCCTCGATGGCTCAGGCTC  
GCGGCCGCGGTACGTCTGCACCTGCTGACCCTGAACCCGCCCCCGGATGCCTCGGCGATCAT  
20 CGCTCACAGCTACTACAGCCCGATGGCCTTGGCTCATGCCTTGGGCGCCACGAGATCGCGCC

5

87

10

15

20

25

30

35

40

45

50

55

TGTCTCGATCACCGTCGTACCGCCGGGGTCGTGCGCGTTCGCGGACGAAGCGATTTCGCGAGCC  
GCTGCAGGCGCTGATCGTGGGCCCCGTGCCTGGTCATCCCCGAGGAGTTTCCCGGGCTCAGCGT  
TCGGCTGCTGGACGTCAACGTCGACGATCCGGCACCGCGTCTGGCGGAGCGGCTCGTGGCCGA  
GCTCTCGGGCAGGATCACATGGTGGCGCTGCGGGCGGCGAGCGCCTAGTGGCCGATGTCGA  
5 TCAAGTCGATGGCCTCGGTGTGGGGATCGCCAAGGTGCCCTTGCGCCGCGAGGGCCACTACCT  
GATTCTCGGCGGCTGGGCGATATCGGCTACCACTGTGCCCCGTATCTGGCCCAAACCTACCG  
CGCCAAGCTGACGCTGACCGCGCGTTCGTCACTCCCGCCGCGCGCTCGTGGGAGCGAATGCT  
GCGCGAGGGAAACCTGGATTCCCGGCAGCGCACGCGCATCGAGCGCGTGTGTGCTAGAGGC  
GTGCGGGGCCGAAGTCCAGACGGCTGCGGTGCACTTGGGCGATCGCCATCGCTTGGCCGATGT  
10 GTTCCGCGAAGCACGGGGCCGATTGCGGCCCATCGCGGGCGTGATTCACTCGGCGGGGATTCC  
GGGACACGTCCACTCGATCGACGAGTGGTGCCTCGCGACGAAGCCCAATTACCGCGAA  
GGTTCGAGGGCTGCACCACTGGCCGAGGTGCTCGATCCGCTGAACCTCGACTTTTGTCTGCT  
GTTCTCCTCGCTCTCGACCGTCTCGGCGGGCTCGGCTACGGCGCCTATGCAGCGGCCAACGC  
CTACATGGACAGCTTCGCCCCGCCACGATCGGCCGGACGAATGTCGTTGGATCGCGGTCAA  
15 CTGGGACGCCTGGCTGTTTGAAGCAAGACGTGCTCGGTGCGCGCCGAATTGGCGCGCCTGGC  
GATCGTGCCCCGAGGACGCTCCGGCCCTGTTGCGCGGGTGCTAGAGCGACTTCCGCAATCGTT  
CATCGTGTCCACCGCCGACCTTCGGGCCCCGATCGACACTTGGATCCGGGACAAGAACCAGCGT  
CCCGCCCCGCGAGATCCGAGCGGTTCAACCGCGACCGGACCTGAGCCAGGCGTACGCCCCGCC  
GATCGGCCCCGTGGAGATTCAACTCTGCGGGCTGGTCTCCGCCTATTGCCGTTTCGACCGGAT  
20 CGGGCGGGACGATTCTTCTTGAATCGGCCCTCAGCTCGTTGACTTGATCCAGCTCAGCTC  
GCGCATTACCGCATCACCGGCAAGGATCTCAATACGACCCAACTGTTTCACTACCCACCGT  
GCGCGCCTTGGCGCTCTTCTCGGCGGCAACCGAGGGGCTCGGCGCGGAGGAGCCCGCCAT  
GGAGAACCTGTGGCTGCAACGAAGCGATGCGACCTCGATGACTGAGACCGAGGTGCGCGACT  
GCGGCGCTACCGACCGCGTTCGAGGATTTCCGCGCAGCGATCCGGGACGACTCGCTGAAGAA  
25 GCGCGATAGAAGAACGAATCGTGTATGAAATACGAAACCACCGGATTGGAATTGGCCGTCAT  
CGGTCTCGCTTGCCGCTTTCCAGGCTCACCCGATCCCGAACAGTTCTGGTCAATCTGCGCGC  
AGGTGCTCCGGAATCCGCCATTTTACGCGATGCCGAGCTGAGCCACATCCCGCATCCCTGCG  
TCACCATCCGCATTACGTCAAGGCCAAAGGCGCGCTGGACCACGCCGATTTTGAACAGCCTT  
CTTCGGCTACTCGCCCAAAGAGGCCGAGGTGATGGACCTCAATTCCGGCTGCTCCATGAGTG  
30 CTGCTGGGAGGCGCTGGAGTCAGCGGCTATGCGCCGAGCCAATTTCGCGGGTCGGATCGGCTT



5

88

10

5

15

20

10

25

15

30

35

40

45

50

55

GTTTCGCGGCGGCGGCTTCAACGACGGATGGATCGCCGGTACCCTCGACCGGCTGCGCACCGG  
CGTGGGTTTGAGCTCCCTGGAACCGCGTTCTTGACCCTGCGCGATTACCTGACCACCCAGAT  
CTCCTATCGGCTCGATCTGCGGGGCCCCAGCCTGCTTGTCCAAACCGCCTGCTCGTCGTCGCT  
GGTGGCGGTCCAGCTCGCCAGCAGGCGCTGATCTCCGGCGAATGCGCCCTGGCCTTGCTGG  
CGGCGTGTGCGCGACCGATCCGCTGCATTGCGGATACCTCTATGAACCCGGCAACATCTACGC  
GCGCGACGGCGTCTGCCGACCGTTTCGACGAGGCAGGCGCCGGTACGGTCTTCGGCGACGGGTG  
CGGCATGGTCTGCTCAAGCGGCTGAGCGACGCCCAGCGGACGCGGATACGATCTGGGCGGT  
CATTCGCGGGGCGGGCGTGAACAACGACGGGCACCACAAGGTTGGCTACACGGCTCTTGGCAC  
GAGGGGCCAGGTGGCTTTGCTTAAAGTGTTTATCGCGCGAGCCGGGTGACCCGGCGACGCT  
CGGCTACCTGGAGGCCCCATGGCACCGGCACCGCGCTCGGCGATCCAATCGAGGTCGAGGCGCT  
TACCCAGGCCTTCGCCAGCAAACGTGCGGCGACCTGCGGCTTGGGCTCGGTCAAGGGCAACCT  
GGGTCACTCAACACGGCGGCCGGCATCGCTGGACTGATCAAGGTGGTGTGCTGGCGCTGAAACA  
TCGCGAAGTGCCACCCACCTCAATCTGCGCCGTCCCAATCCGAAAATCCGCTTCGACGAGAC  
GCCGTTTTTCCAGTCGTGAGTTGCAACCTTGCCAAAGCGGGACCGGCCCTTGCGAGCCGG  
CGTGAGCTCCTTCGGCATCGGCGGTACGAACGCCCACGTATCTCGAGGAGGCACCGCCGAC  
GGCCAACCCGGCGCCACACGGCAGATTCCGACTGTTGCCGCTTTCGGCCAAGACACCGGCTGC  
GCTCGAAGCGAAGCGCCGCGATCTGGCCGGCTTCTCGAACGCCACCCGGAGACCTCCTTGGC  
CGACCTCGCCTTTACCTGCAACGCGGCCGCGAGGTCTTCAGTCACCGCGCCTGCCTCGCCGT  
GGAGACCTTAACGTCCGCGCGCACGCGGCTGAGCGGCGAGTCGTCGAGCACTTGGCTGGTGGG  
CCCCGCGCCAGCGCCATATTTCTGTTCCCTGGTCAAGGCAGCCAGCTCGCCGGGATGGGCCG  
CGGTCTGTATCACCATTTTCGAGCCGTTCCGCACGGCCGTCGATGCCTGTCTGCGCGAGCTGGA  
GCCAGGACTGCGGCAAGCGCTCAGCGCCCATTTTCGATCCGAATCGCGGCGCGGACCCACCCGA  
TTCGACGACCTTCGTCCAACCCTTGTTGTTCCCTCGTCGAGTACGGGGTGACCGAGTGGCTACG  
CTGCTTGGGTGTGCGGCCAACAATGGTGTGGGTACAGCTCTGGCGAGTATGCCGCAGCCTG  
CGTCGCGGGCGTTCTGTGCGCGTCCGCGGCGGTCTCGCTGCTGGCCGAGCGCGAGCGGCTGCT  
GCGCGACCTGCCAGCCGGCGCCATGCTCGGCGTCCCGCTGGCCGCGGAGGCGCTCGAGGCGAT  
GTTGCCCGACGCTCTCGATCTGGCGGCGATCAACGGCTGTCAGCTTTGCGCCGTGTCCGGGCC  
GGTCGCGGCGGTCCACGCCTTCAAGGCCCAACTGGAAGCCGCCGACATCACGCCCGCCTGTT  
GCACACCGATCGCGCCTTCCACTCGCGGCTGGTAGCACCGGTGCTTGACCGGTTCCAGGCAGC  
CGTTCAACACGTGGAGCTGCGGCGGCCGCAAGTACCTTACCTCTCGACCGTCAGCGGGCGATT

5

89

10

5

15

20

10

25

15

30

35

40

25

45

30

50

55

GGAGGCGGATGGGCCGGCGAACCCGCACTACTGGGTGCGTCACCTGCGCGACACGGTGCGGTT  
TGGTCCAGCCCTGGAGGCGCTGCCGCCGGTGGATTCTTCGTGTGCATCGAGGTGGGACCAGG  
CTCGGCCTTGAGCACCATGGCGCGCGAAACGTTGGGTTCCCAGGCGCGACTGATTTCTGTGCT  
GCCGCGGCCGCGAACGGGGCAAATCGAGCCCGGTCCGGTATTCGAACGACTGGCGGCGCTTTG  
GCGCAGCGGGTTGACATTGGATTGGTCTAAATTGACGGGCGGCGAAGAGGGTCATCGAATTCC  
CTTGCCAGTCTACCCGTTTCAGCGCAGCCATCTGTGAGCTCCCTGGCGGCGGGCCACACGCC  
TTCGTGCGGGCCTGCAGTCGAATCAGGCGCCATCCTTGCCGAGCGATCCGACGGGAAAAACGC  
TGAAACCCGGGATTGCCCGCTGCCAACCGCCACGCTCGAGCCCAAGGCGGTGCTCCGGCCCC  
ACTCGAGGCTACCGACGCGCGCAGGTA CTGCGAGCGACTGGCCGAAC TTTGGCGCGAGTTGCT  
AGGGTTGACCTCGATTGGGCCCGACGACCATTTCTTCGACCTGGGCGGCCACTCGCTGACCGC  
CACGCGGCTGCGCGCCCTGATTACACAGCGGTTCCATGTGATCTCGGGCTCGACGAAATCTT  
CGCTCATTCGCGTCTCTCCAGCTGGCGCGCCGTATCGAGGCGGCGCCAAGAGCCGATTTTC  
CTCCATTCCCAGCGCGCCGGACCAGGACGACTATCCCTTGTCATCCGCCAGCAGCGGATTCA  
CAGCATCGTCACGAGGGCCGAGGTCCGCACTGCTTATAATTTCCGATCGTCTCTGAGCTGCA  
GGGCGCTCTGGATCGAGTGCGATTGAGGCGACGTTCCGCGCATGTTCGGCGTCATGAGGG  
GTTCCGCACCCGCTTTGTGATGCGCGATGGCGGGCCGCGCCAGCGCATTGTACCGGACGTGGC  
GTTTCGCTGCGCTCACCCAGGTGAGCCAGAGCAGGTTCCCGGGCGCATCGAGGCCTTCAT  
CCGTCCCTTCGATTTGGAACGCGCGCCGCTGTTCCGCGCGGAGCTGTTGAGTTGGCCGAGCA  
GCGCCATCTGCTACTTTTCGACATGCACAACTTAATTGCCGACGGTATCTCGCTCAACCTGTT  
CGTCGCCGATTTGCGGGCCCTGTACCATGGTCGTCCGCTGGCGCCGCTGAAACTCCGCTATCG  
CGACTATGCCGTTTGGCAAGAGGCGCGGCTGGCCTCCGATGACCTGCGCAGCCAGCGCGAATG  
GTGGCACCGGCGGCTTTGCGCCCGGTGCGCCACGCTGGCGCTCCCTCCCGATTTCGCGGTCC  
GGCGGTGCGCCGCTACAAGGGCCGTAATGTGGTGTTCCACCTGGACCGGGAGATCCGCGACCG  
CCTGGTGGCCCTGGCTCGAACCAGGGGGTACCATGAACGTGATGATGCTGGCGCTCTGGGC  
TGCGCTGCTGCATCGCGAAACCGCCAAATCGGAGCTGGTGGTCCGATCGCTGCTCGGCGGGCG  
GCCGCACAGCGAGCTGCATCCCGTGATCGGGCTCTTCACCAACTTTTGGCCCTGCGGTTGGC  
GGTCGAGGGATCGACCCGCTTCGATCGCTTCCTTGCCGCTTGCCACCAGGTGTTTCTCGAAGC  
CTATCAGCGCCAGGACTATCCGTTCCACTTGTTAGTCCAGGAACTCGTGCCGGTCAGGGACCC  
GTCGCGGTGCGCGCTGTTCCAGACCTCGCTCGTCTACCACAACGAAATTGACGGCAAGACCAA  
GCTGGAATTGGAAGGGCTGAAAGTCGAAGTGTTCCCTTCGAAAAGGTGTGGCGAGGCTGGA

5

90

10

15

20

25

30

35

40

45

50

55

TTTGAAGCTGGATGTGACACCTTTTTCCGACCGACTCGAATGTGTTTTGCAATACGACTTGGA  
TCTGTTCTGCGAGGAGACGATGCGCGGCCTGATCGCGCGGTTCAGGCGTTGGTGGCGGGGCT  
TGTCGCGCGATCCGGCGCAATCGCTCGCCGCCGCGAGCGTTTCCGGGAAGCGGGCGCTGCGCGC  
GGGCGTGGCCACGGCAAGCGAATCGTCGCCGAGTCACTGCCGCCGAACCATCGACGGCGTA  
5 CGCCACTCCCTCACCAGTCAACCGTCGCGGCTAGTCTGACGGGACCCGCCGACCTGCCCGC  
15 GATCTTGGCGGCCTACGTGGGGCAGAACCCCCATCCGTTGCGGATCCATCGGGGTCTCATTTT  
GGAGGCGCCGCTGGGGTTGCGAGCGCTGCGGTGCGCGCTGGACGAGTGTCTGGAGAACAACAC  
CCATTGGCGCAGCGTGCCTGCGGGCGATCGCGCGCGGCGGTGGATAAGTTGGAATTGACCAG  
CCTGGTGGCGCTCGACGACCTGCGCGGGTTGGTCAATCCTCAGGCGAATGCCTTCACCCTGGC  
20 TTGGCGCGATCTGGCGATGCCGTTCCGGGAGGGGCGTCCCCTGTGGCGACTCCGCTGGCGTG  
10 GTCGGTCCATCGCGCTGGTTGCTATTGCTGACGTTTCAATCCATTGATCGGCGACAACGGCAC  
GGTCGACCTCTTTCTGGCGGCACTCGCCGATCACCTGCGCGCGCGTCCGCTTTTCCCGTAGC  
25 ACCGCTCGATGAGGCCGAGCTGGAGGCGGAGCTGAAGTGGGGAGAGGAAGGGGAGGGCCTCGG  
GCTGACCGCGATCGCGCCGGTCTGGGCCAATTGCGCGAAAGTCCGCTGAGTCTGTGGCCCA  
15 GATGTGGCTGGACGAGGTCTGTGCGCGCCACGACCTCACCCCGCTAGAGGTCTTGGCGGCCCCG  
GCTCCTCGATTGGACACGAAGCCACGGTCACGGGTGATCGCTTTGTGGACGCCGCTGCCCGA  
30 GGACCATCCGCTTCGCGATGAAGGCCGCTGCCCTCCAGGTTGCGCTGCTGGAGGGGCCCGCTC  
GCAGCGAGGAGCGGGCGATCCAAGCTGGCTCGAGCAAATCGCCTTGAGACGGGGTACCCCTGC  
AACGGAGGTGCTTTGCCCTACTCCGACCCAACGGGCAGCCATCGACCTCGCGCTGGCCTGGCT  
20 GCGCGAGCCGCTCTTACGGTTTGGTTCGGAACCGTTCAGCCGTGGCCGGAATCTCCATTGGT  
35 CTGTCCGTTTCCCTCAATCTCGCGTTCCGGCCAAGCCATCCAATTGCCTACGCGCTCAAGCA  
CGAGGCCACGCTCGCGGTACGGCACGGGCGCGGATCTGATGCGTTTCTCGACGGCTTGGG  
CCCGAAAGCTGAAGATTAGCATAAGCGCCCGGCCAAGGGCATCCTAGGATGACGCAAGCCTC  
40 GGCCGCGTCGACGTCCCAGGTGCGCGCGGAGGTCACCCCGGCCGAAAGACGACGATGACGA  
25 TCAAATCCGAGATGTCGGCCGTTGCTCACTCTGCGGAGAGCGGCTTCCGCGCTGGGCCACGCG  
TGGGCGGCGCGATGAAGCGGGGCCGACGCCGAGCAGGCCGCGTGAAGCTGCTCCGCGCCC  
CGGTGAAGCGGAAGTGGCTGCCCCCGGCGCCCTGCGCCTGAGCGAGCGGCGTATCCCGG  
45 AGGTGTGGGCAGGCTACCGCGGAGCGCGGGATGACCCGAGCCCCGCCCGCGCGGACCAT  
GACGCCGCCCCACGGGGCGAGTCGTCCGGCGCGCGGCGCGCTCGGGGCTTCCGCCCGGGG  
30 CGGGCAGGTGCAGGATGGTCCGGCATGGTGACGCTCCGACGTCCGACGGCATCGAGGACGAG

5

91

10

15

20

25

30

35

40

45

50

55

CTCGCGCCGTTCCCCCGGTCTGCGCGGCTGGCTCATCGAGGGCGAGCTCGGCCGCGCGGG  
ATGGGGCGGGTGTTCGGGCGCGGCACCCGAAGACGCGGGCGGGGCGGCGATCAAGGTGCTG  
CTCGGCGACTACGCCCCGCGGCGGACGTGGTGGCCCCGCTTCGGGCAGGAGGCGATCGCCGTC  
AACATCATCAACCACCCGGGAATCGTCCGCGTCTTCGACTCCGGCGAGCTCGAGGACGGCTCG  
5 CCCTACATCGTGATGGAGTACCTGGACGGCCGGGGGCTGCGCGACTGGGTGCAGGCCGTGCCG  
CCCCGCGAGCGGCCGCGGCAGGTCTGTGCGGCTCGGCTACCAGATCGCCTCGGCCATGGCCGCG  
GCGCACGCGTCCAAGGTCTGTCACCCGATCTGAAGCCGGAGAACATCATGGTGGTCCAGGAC  
GAGCTCGCGCCCCGGGGCAGCCGCGTCAAGATCCTCGATTTCCGGCATCGCGAAGGTCTCTGG  
GGAGGTCTGCCCGAGGTGCTGGAGCTCGAGGGGAGAGGCTCCCTCGCGCCCGCGTCCGCGTCC  
20 10 ACGATCCGCACCGAGCTCTCGACGCGGCCGCGCCGACGGTGGGCGCCACGACCGGCCAGAG  
AGCCCGCTGGGCGCGAGCGCCACGCCAGAGAGCGCCCTGGGCGCGAGCGCCACGCCAGAGAGC  
GCCCTGGGCGCGAGCGCCACGCCAGAGAGCGAGGCCACGAGGAAGACGCGCTCCGGAGCCTC  
CCCGTCTGTACCAGCGGCAGGCCCGCGATCCACCCGCGCCGGTCTGAGATCCCGCCCGAGGCG  
25 GTCTCTCCGCGGCGTCTGCGCGGGTCTGCGCGCGTCTGATCGAGCCAGGCGCGCCCGCGCCGAG  
15 AGCGAGGGCGCGGACAGCCCACGATGCCGTTACGCAAGAGGGCGTGTGGGGCTCGGGACG  
AGGAGCTACATGGCGCCGAGCAGGAGCGCCACTCCGGGAGCCTGGACGTGAAGGCGGATGTC  
30 TACTCGCTCGGCGTCATCTCTATGAGCTGCTCGAGGGGCGGACGCCCCGACGCGCCGAGCGCC  
GCGTGGCCGCCCCGATGAGCGCCGCCACGCCGCCGATCTCGTCCGCCCTCGTCCACCGGTT  
CTGGCGTTCGATCCCGATGCGCGGCCGCGCATGGCGGAGGTGGCGAGCGCGCTTACCGGCTC  
20 GGCCGGGCGAAGAAGGAGCTCGACGAGGCGCTCTCGAGGTGGGTCTGCGCGGAGGGGCGCCG  
35 GGCTCTTGCCGTGCGGCTATGCTCTTCTCGAACTGGTCCTCCTGGGCCCTGGGAATTATAC  
GATTCTTTCCAGCCTGTAAGTGCATTTTTCTTTCAATATCGTCTCTCTTCATATACGAGGTG  
AGTTCTCTGAGGTCTCTATAAGTCTGGGGTGTCTATTCCGGCCTCTTACTTGTTACTTCGC  
40 CTTCTTAGGAGTTTTTCTTAATTTTGCCCTCTTACATTCCCGTATTCAATTCTAACTGGGCC  
25 TATCTCATTCGCTAATACGTTTCTGTATTGTGTACATCTCCTATCATGTGTCAATACTTGTTT  
CTGTTTATCATTATTCTTATTGTTTACGCTCTTATTTCAATCATAGTATAACATTAGTTTACT  
GATTATCGCACTTGAATTCGCG

or its complementary strand,

30

5

92

10

(b) DNA-sequences which hybridise under stringent conditions to regions of DNA-sequences according to (a) encoding proteins or to fragments of said DNA-sequences,

15

5 (c) DNA-sequences which hybridise to the DNA-sequences according to (a) and (b) because of a degeneration of the genetic code,

20

10 (d) allele variations and mutants resulting by substitution, insertion or deletion of nucleotides or inversion of nucleotide segments of DNA-sequences according to (a) to (c), wherein the variations and mutants offer isofunctional expression products.

25

8. DNA sequence according to claim 6 selected from the following

30

(a) open reading frames:

		<u>Nucleotide Position</u>		
35	20	ORF1	1666 - 1	Seq ID No 3
		ORF2	1605 - 3338	Seq ID No 4
		ORF3	6100 - 3398	Seq ID No 5
		ORF4	7110 - 6374	Seq ID No 6
40		ORF5	9590 - 8433	Seq ID No 7
	25	ORF6	11393 - 9855	Seq ID No 8
		ORF7	13656 - 12712	Seq ID No 9
		ORF8	15374 - 18984	Seq ID No 10
45		ORF9	20003 - 27889	Seq ID No 11
		ORF10	28251 - 29402	Seq ID No 12
	30	ORF11	31720 - 30401	Seq ID No 13

50

55

5

93

10

ORF12	31982	-	32932	Seq ID No 14
ORF13	33128	-	33613	Seq ID No 15
ORF14	33661	-	34007	Seq ID No 16
ORF15	35611	-	35255	Seq ID No 17
5 ORF16	37856	-	35730	Seq ID No 18

15

or DNA sequences complementary to said open reading frames,

20

(b) DNA-sequences which hybridise under stringent conditions to regions of DNA sequences according to (a) encoding proteins or to fragments of said DNA sequences,

25

(c) DNA-sequences which hybridise to the DNA-sequences according to (a) and (b) because of a degeneration of the genetic code,

30

(d) allele variations and mutants resulting by substitution, insertion or deletion of nucleotides or inversion of nucleotide segments of DNA-sequences according to (a) to (c), wherein the variations and mutants offer isofunctional expression products,

35

and peptide sequences corresponding to said open reading frames

40

SEQ ID No 19 (>ORF1)

25 VDPEREAVTLGLAFNRAQGRTYARGPEARA EYIGTAMRAADVIEDRFEIERLAVSGGMGDVYR  
ARDRVSGQAVALKVLQGASANDLRRFA REAEALVTLRLPGVVQYVAHGVTGAGRPYLAMEWLD  
GVTL EERLAGAPLTLAESVALAARVATT LGAIHWLGVVHRDLKPSNMLVGGAVERTLLDFG  
45 IARHLRLAPTLTSPGAVLGTGPIA EQVRGDAPVDARDVFALGCVLFQCLAGRPPFLGNSAL  
ALLMRVVLEEP PRLGELRDGIPEPLERLVARMLAKNAGERPRDGAAAAAE LAAVAGEGLSIGA  
30 SAVAAPAAPGEAITTAERKVMCVILAEDGGAEAGATLSEDDGAARAEALRDIAARHGGRDLRL

50

55

5

94

10

QARWWLVALSGAESPTDLATRAAHCALALRAALGGVPVSVATGLAEVEARLPVGELVDRVAQL  
IAGRDLGSPPEIRLDDATASLLASRFETVQGGGCWLRGPKEEPDAVPRLLGKPTPCVGRERE  
LSQLATEWRHCVDEPSANAVVVVGAPGLGKSRLAWEFLLRTLEQREGAAI

15

5 SEQ ID No 20 (&gt;ORF2)

20

VRPCARLNASPSVTASRSGSTAAGSVHASTSACVEQPATGRTQPASPRWPPGAAALRLTSAMP  
RWFNTAGPCNPADHYMLPAEERLPAVRDLVDRKAYFVLHAPRQIGKTTSLRTLAQDLTAEGRY  
VAVLVSAEVLGAPFSDDPGAAELAMLAEWRTAGAQLPADLRPPPPFDAPAGQRIGAAALRAWAQ  
AAPRPLVVFLDEADALRDATLVSLLRQIRSGYPDRPRDFPHALALVGLRDVRDYKVASVDSGR  
10 LGTSSPFNIKVESLTLRNFTREVAITLYAQHTAETGQVFRPDAVDRAFELTQGGPWLANALAR  
QLVEVLVKDRAQPITSANVDRAKEILIERQDTHLDSLVDRLREPRIRAVIEPMLAGTALPSVP  
PDDLRFALDGLVRMTAEGGLDVANPIYREIIVRELAFPIRASLPQIKATWLTQDGRILDADRL  
25 LDAFLSFWRQHGEPLLGAAPYHEIAPHLVVMAFLHRVVNGGGTVEREYAIGRGRMDLCVRYAG  
ETLAIELKVWRDGRPDPAEGLAQLDGYLAGLGLDRGWLILFDQSRGQPPIAERTRRERALS  
15 AGREVAVIRA

30

SEQ ID No 21 (&gt;ORF3)

35

VTIKKTFRSIDPATLPKHFDSPVAELRLADLWEADGTYRYDPSRPREETFVVDTPPPTASGSL  
HIGHVFSYTHTDVVVRQRRMRGNIFYPMGWDDNGLPTERRVQNYFHVRTDVRTPYERGLTLP  
20 QAAPETIKKEPPRIVSRPNFIELCHKVTREREQVFKALFRRVGLSVDWRNEYATIDDHCRRTA  
QLSFLDLHEKGHLYSVFAPTMWDVDFQTAVAQAEVEDRPQSGAFHDIAFAVEGTAEELVIATT  
RPELLAACVGVTAHPEDPRYQHLFGKTALTPIFRAPVPIFPSPLVDREKGTGILMVCTFGDAT  
DVIWWREQKLPLRQMLGKNGRVLPVTFGEGAWESRDPAANAAYAPLQGRGVKQARAAVVELL  
40 RREEHAAAPGRGPALRGEPRPIERAVKFYERGDQPLEFVPTQWVRLADKKAELLEYGDKIK  
25 WHPDFMRLRYRNWTEGLQGDWCISRQRYFGVQFPVWYPLDAEGNPDHSRPLLATREMLPVDPT  
VDVPPGYEASQRDQPGGFTAESDVFDTWFTSSLTPISSHWGDDPARHARLFPADLRPQANDI  
IRTWAFYTIKAMLHESSVPWHHVAISGWILDPDRKKMSKSGNVVTPMHLLDTYSSDAVRYW  
45 SASARLGTDTAFDEKVLKIGKRLVTKIWNASKYVLSQSAEVHPISEELDRALLHKLSAVVDDA  
TRSFDEHEFAAALERTEDFFWRWFTDAYLELAKARARGEAGEAARGSAVAALRLGLSVLLR  
30 LFAPVLPYITDEVWRWVYAEETGDTSIHRAKWPSAADFAAVAAPSDPGLLDLAAAAAAMAVNKR

50

55

5

95

10

KSELGASVGRVVTDLALGANAATLARKPALGDVLTAVRAGAHALVRPELADGEVLVVRCELE  
PAAAAAAGAGGAAASEE

15

SEQ ID No 22 (>ORF4)

5 MIHAEPFEARLVAAARPLSPFVRELSFERADGRSFLFEAGQWVNLVLPPLPGGEVKRAYSIASAP  
DGSPRFDLAVTLVQGGAGSEHLHRLPEGATLRAIGPHGLFTRDPGDSAPSLFVATGTGITPLR  
SMLRASLRAGLAAPHLWILFGARFEEDVIYRDELEALARGSDRIRYEITLSRGGPSWAGRRGY  
VQAHVPELYRELAEKSGDPAPHVFICGLDRMVSSVRELARGELGVHRKHVHVERYD

20

10 SEQ ID No 23 (>ORF5)

MKSLPSDRAARLAQSDIRMTLACAXVHGINMSQGVCDTPVPSVILQAVKEAMDRGCNTYSRF  
DGIVELRHAIAAKLARHNGIAADPETDITVSAGATGAFQATCMALLNPGDEVLLFEPFYAYHA  
25 QAILAVEAVPRYVTARSLSWNVGDDELERAITPKTKAIVVNSPGNPSGKVFGRMELEQIADLA  
CHHDLMVITDEIYEYFIFDGREHVSVASLPRMSERTITIGGYSKTFSTITGWRIGYSVADARWA  
15 KAIGAMSDLLYVCAPTPLQHGVAAGIRGLPRSFTYGLAQGYERKRDRFCRALEKAGLPFCVPQ  
GTYVVLADVSRLPGRTGRERAIYLLDETGVAGVPGDAFFEGTQGSRFMRFCFAKTDEDLEEAC  
30 QRIEQLA

30

SEQ ID No 24 (>ORF6)

20 VSDPRKERLGMDLEEFRRIGMRIIDWAADYLGHDPDRYPVFPAIRPGDVKGRLAPTPPVEPEP  
MDAVLTDFEQIILPGITHWNHPRFFAYFANTASGPGILGELLAACLNVNMLWRTSPAATELE  
35 ELVLSWLRQMLDL DAGLHGAIMDTASTASMVAIAAARSDAEP TIRLRGMAGQRRMRLYASEQA  
HSSIEKAAITLGIGQEGVRKIPTDFAFRMVPEALRAAVVEDLGAGLRPFCAATVGTTSTTSV  
40 DPIP AIVSVCREHGLWLHVDAAYAGMAAIVPEHRDVLAGEGADSLVNVPHKWLFTPMDCSVL  
25 YVRDADRLKRAFSLVPEYLRTEGDVTNYMDWGIQLGRRFRALKLWMIVRYFGHEGLAARIREH  
LRIGQQLAQWVDADPDWERLAPTPFSTVCFMRPSALACIMRSADEAERESIERELDRLEAL  
LDEVNKSGRVFLSHTRLHGRTYIRVAIGNIRSDEVAVREAWECLEAAGARLCADERFVSCSR  
45 ADEGRGKS

45

50

55



5

96

SEQ ID No 25 (&gt;ORF7)

10

5

15

MRREEPVLEAFYERYCAAPRETSYHVELPVDVELHQEAAPALPOARSLELAGRVALVTGSSRG  
IGKAIALRLAEQGADVAVNYHSNKDAAEQTAAEIRALGRRTMVVQADVTRPNAAAELFSSVEA  
QLGPIDILVNNVGDDFFKPLAAMTDDDEWRNVMDSNLSSVHYLCRAAVARMRQRKSGRIINIGL  
SPTYAIRGAPNVAAYSIAKTGVLILTRSLATEEAPHGILVNCVSPGLIDNGYLPQAQKEWMER  
RVPMGRLGRASEVADAVAFLASDRASYVSGANIAVAGGWDWTDGRGTEHRRVDLFIGHEEP

SEQ ID No 26 (&gt;ORF8)

20

10

25

15

30

35

40

25

45

30

50

55

MSGRFPGARNVEELWQKL RAGVECVVTFTEAEALAAGVSREMLANPSYVRRGAPLDGVELFDA  
SFFGFSPREAESMDPQQRIFLEVAWEALERAGYDPDAHSGPIGVFAGSAPSGYHSLAQSDPEI  
LGALGHYQLTLNNDKDYLTTHASYKLNLRGPSVCVQTSCTSLVAVVMACQSLLNHECDMALA  
GGVGIIHAHQRRGYLYQENGISSPDGHCRAFDVAAKGTVGSGIGIVVLKRLADALADGDHVHA  
VIRGAAINNDGSSKIGYTAPSVQGQAEVIGMAQALAGVEPDDISYIEAHGTGTPLGDPPIEIAA  
LTRVPRAKTARRQFCAIGSLKTNLGHLDAAAGVASLIKTVMALEHRELPPSLHFERPNPKLEL  
ESSPFYVNTRLTPWHAARGPRRAGVSSFGIGGTNAHVLEEAPAPPPSGPSRRWQLTLAARS  
EAGLARATADMI EHLDRHSGTSIADVITYTSHVGRRAWPFRRVVGESAADLRAALASEGSPRS  
ISSCQAARERPVVFLFPQGQAQHLFMARELYEVEPIFRQSLDRCAELLRGPLGLDLRQVLYPA  
EGQRDDAEQELGRTAIAQPALFAIELSLAKLWMAWGIVPQAMIGHSVGEFAAACLAGIFREED  
ALRLVAERGRLMQMPPGAMLAVPLAEPAPYLSDDISLAAINGPALS VVAGPIEAIDALAA  
ELLDHGLSCRR LHTRHAFH SKMMA PVVDAFTRCVSAVERRPPSGHFLSTLTGGWISPEAATIP  
AYWARQLVEPVRFQAQAVRQLLSESTWLWLELPGQQLSPLVRQQARADGGQVVVASLPRAKDA  
GADHLAVIEALGRVWSAGGTVDWKRFEHEGEARRRVLLPTYPFERQRYWASPRHTSAPPEAIK  
PLLAKNPVADWFFLPARRSDPPVSFDAQAVTTRRSTWLVFIGDEGLGAALVEGLARRGHEV  
VAVVTGERFEQTGTQRYTIDPAANGDVASLFARLEIEGRMPDRI VHAFCTSPADGARIERGAA  
LEIERRLGFD SLLLLAQVIAAQRHPKPLMLGVITTRAHSVIGTEIEPLRALVLGPCRVIPQE  
IPHVSCRNIDIDLPGEGGRAEIAARLIADLERESPDSVVAYRGRRWVESIELTDVGRRSAGA  
APRLRQRGAYLITGGLGGIGLVAAELLAREAHARLILVGR TGLPARQGWDDWLAAHGAGDATS  
RKILRIRALEEAGAEVKIAAADVSDFNAMRSVIEEARTFRIGRIDGVIHSAGIASGGMIQLRTP  
MAAHRVMA PKVGGTLVLDALLRDERPD FLLICSSLASLVGGATQIDYCAANAFDAYAQSREG  
EEGCRVISVQWDTWSDVGMVDFKLPADLQEGRRRESLKRGISSEGAEVLGRILSAGMSGPLA

5

97

10

ICTSDLPAYKQSVTTTTRSQHEQTPAARPMHSRPTTTGAYVAPETETERRIAAIWQDLLGLEQV  
CANDDFLQLGGHSLLATQVLSRVLQTLKVGISLPQFFDAPTVAGLSRLVDAARAEGAGPVAPA  
IGRVERDAYRIKPPAAEQAARTKP

15

5 SEQ ID No 27 (&gt;ORF9)

20

10

25

30

35

40

45

50

55

30

MEPVGGVDMNQPAKQQETCVFPTSFAQRRLWFLDQLEPGSAVYNMPASFRTTRGPYDVDSLVR  
VNEIVRRHESLRTTVDIVIDGEPVQVIAPSLRIEVPVVDLSEIDEPERAEARRLMAEESRRPF  
DLTRGPLLRKALLRLGEADHVLILTMHHIVSDGWSMDVLFKELSTLYAAFHEGRPSPLPELPI  
QYADFAVWQRELLQGEVLESHLGYWREHLRGAPTLLLELPMDRPRPPAQTFRGSQRAFRPLPSL  
QQAVQALSREQGATPFMTLLTAFSVLLSRYARQSDLVVGTPIANRTAELEGLIGFFVNMLAL  
RIDLGGDPSFRELLGRVREVTLGAYAHQDLPPERLVEELSPGRSPSHSPLFQVSFTLQNTPMD  
ATNRADIASGGAPLVEMKAAKFDLILELSESPQGLLGTFEYNTDLFDAGTIERMAGHLEVLLS  
SAVAAPDRPIAELPLMGAERSRVLVEWNSTAALYPEDHCMHELFEQQVERSPEATAVLLQQQ  
TLTYRELNMRANQLAHLRLSLGVGPEVRVGLYLSIETVVAAILGVKAGGAYVPLDPTYPSE  
RLGLMMADAAPSVLLTQASLLSKLPHPGDATLVQLDALHEALSRLPHHTPRSGVTAQNLAYVM  
YTSGSTGRPKGVLEHRLCNLPTVQAKLYGIAPGDRLLQFAPLCFDTSFCEIALALLSGATL  
VMGTADELLPGPPLVELLKKHAVTAMLLAPTTLAALPEQQSAALPLRVLTMAEACPAELVKR  
WKAPGRRLFNISYGPTETTIWASSAADLSDERIPPIGRPIANTQIYVLDEALEPVPIGVPGEIF  
IGGVGVARGYHGRPDLTAEFVDPFGQTKGARLYRTGDRARWLPDGNLEFLGRNDEQVKVRG  
VRIELEEIIRAALLKHPAVAQAVVREDTPGDKRLVAYVVGRGGARVTAAELRQSVSERLPAT  
MVPSSFVALDALPLTPNGKVDRLALPEPEQSAGGEDHVAPRNAVEEELARIWASVLRRLERVGV  
HDNFFEIGGDSILSIQIVVRAQQAGLRLTPROMFQHQTI AELSTVARAVEAVHVEQDPVTGPA  
PLTPVQRWWLEQEAAPHFNQSI FLEVRERLDESALEQAI AHLIDHHDALRLRLARDERGAH  
QVFAAPGGSTPFQVRDLGALPSAEQISAMEKAASEAQASDLAAGPVVRVAVLFDLGEVAPQRL  
LVIAHHIAVDSVSWRILLDDLFGAYEQARRGEAVRLPPKTTSVKRWAELLTEHAGSEAVKAEL  
GYWLDSSRRTVAPLPVDRRAGEDVWGSARHIVVSLTPEQTEQLLREVPQAYRTRIDDALLTAF  
AQAIARWTGSPAVLLDLEGHGREELAGVDLTRTVGWFTAMYPIILLRVDAADPGEALKSIKEQL  
RAVPGRGLGYGLLRYSRTIAEVRALPQAEELCFNYLGQLDQAIPEAAPFRPAREYQGSERSP  
GAHRAHLIEVNASIANGRLYATWTYSERRHEPETIERVAASFVTALRALIAHCTLPEVGGNTP  
SDFDKVRLRQETIDALDAIDAGPGPSARGSRIEDVYPLSPLQEGILFHTLYATDYTAYVEQFH

5

98

10

15

20

25

30

WTLEGDFDAEAFTRALQDVVARHAAALRTSFAWERLDAPLQIVRTGAVLPVEHQDLRGLAAEEQ  
TAHISRYVEAERQRRFDLRKAPLMRAGLLRLRKDAWCLVETIHHLILDGWSTQILLKEVFPLY  
EAHRGHRCHLALALEEQPRPYGDYIGWLAKQDQVRTAAFWRRELEGFSAPTPLGVDRAPPHDDG  
GPRFGWRRIALSGDDAARLAAAFARQHQLTMSTLVQGAWALLLSRYSGDPDVLFGMTVSGRSAP  
5 IPGIERMTGLFINTIPVRVREPADASVLAWLKALQEHEAEELLEHEHSPLVEVQAHS DVPRGTP  
LFESLVVFENYPVQVIFEAPPVEGPTRAEEGLRMIDAQYISDPPYPLTVVAAFHGTLYLNIGY  
ERRRFDDQAVERMIGHVTLLRGFVQRPETSVRDLPLLTAAEEERTQLHAWNATAAPYPEGHCHM  
HELFEQQVERSPEATAVLLQQQTLTYRELNIRANQLAHHLRSLGVGPEVRVGLCLERSIETVV  
AILGVLKAGGVYVPLDPTYPSERLGLMMDAAPS VLLTQTSLLSKLPPHG DATLVQLDALHEA  
20 10 LSRLPHHTPRSGVTAQNLAYVMYTSGSTGRPKGVLVEHRGLCNLPTVQAKLYAIAPSDRLLQF  
APLCFDTSFCEIALALLSGATLVMGTADELLPGPPLVELLKKHAVTAMLLAPSVLAALPEQQS  
AALPLRVLAMAGEACPAELVKRWKAPGRRLFNSYGPTETTTIWASSAADLSDERIPPIGRPIAN  
TQIYVLDEALEPVPVIGVPGEIFIGGVGVARGYHGRPDLTAEFVDPDPFGQTKGARLYRTGDRA  
RWLPDGNLEFLGRNDEQVKVRGIRIELEEIRAALLKHPAVAQAVAVVREDAPGDKRLVAYVVG  
15 RGGARLTAAELRQSVSERLPATMVPSSFVALDALPLTPNGKVDRRALPEPERSAGGEDHVAPR  
NAIEEELTRIADVLGAKRVGVHDNFFDLGGHSLLLVRVHDLRGQRFRDRPPSMVDLFTYPTVA  
SLARFLGERANGKQSPREAAADVTERGRRRLEARARRAKAIRGPT

35

40

SEQ ID No 28 (>ORF10)  
20 MKHNIGWLLPAALATLAFVPACSPNHGEDAPSVTSAESGAAPSADCVALGAKLQAALDGAAAA  
QKAPGAAAAVQSGDCVWRGATGVSDLVASTPTKPGDLFRIGSITKTFVSTLIIMLRAEGRSL  
DDAVSKYVKGIPAGDQMTLRQILGHTSGLFDYTYSPALGQMIEVDPTRAFAPAELIALATAEA  
PYFAPGAGFRYSNTNYIVAGLVAEAVSGGTLAGLLRTRIILDPVGLAHTYLDGAEPVQGLIRG  
YG DYAGLVDITDQLSPTEAWAAGALVSNVDDLNRFFALLISHELLSSDELQDMTTWTPTMWP  
40 25 HEPGYGLGLIERDSALGSLNGHCGIIWGFQASASYGVPGRGDAITALINRSDGDAARLVDELAK  
VVKER.

45

SEQ ID No 29 (>ORF11)  
MSIDRAVLEQLDRVGGRLAEGKALKLLEDIAWPREVEERFFAAGEDRLPEVEYRVDRDGLARR  
30 VAELRELLCAIDGDAPALCWLNRDQVRAQIQAAELLEAAGTRAFSARSQELYGCARSRFFGGSL

50

55

5

99

10

RNIDLAEHLTERLRVHGWDEASDPEEEPLDAGALRDMLAARVAGRAPRLDLEITVDPRVTAKV  
VAGMSRVIRPEATFAAWEAEGLWHHEVETHALTAHNGAAQPRCAFLRSGGPRTTTRTQEGLAI  
FAELYSRSLSIGRLTRLAERVRLVDMAEQGASFLDLYRHLRERGAERRDAYFDAQRVCRGGLV  
EGGAPFTKDACYLAGLLEVYAFLLAAVLRGGLRDEVLLVCGRIALDDIAVLAEALRAAGVLERP  
5 RYLPGWLRWQTLTPYFAFTSFMDGIDLGPVERHFQELLRVAADARPAGEGRRRRGRPREG

15

20

10

SEQ ID No 30 (>ORF12)  
MSESVAQLEEHRAALTGH CYRMLGSVVDADDAVQETMVRAWRS LDKFDGRSSLRTWLYRIATN  
VCIDLRADRARRARPIEEGPVGTVDDALETRPRTHWLEPVDAHALPADIDAAERAMLRQSIR  
LAFVAALQHLPPKQRAALLTEVLGWSAAEVADSLNTSVAAINSALQRARATLASRDLGDARP  
SLPEPQSALLDRYVNAFER YDVALTALLHQDATLSMPPTLWLRGHESIRAWLVGPGAGCRG  
SRLIPTAASGSPAFAQYRPAPEGCHRAWALIVLDVAGDRIVSMTSFLDTETLFPFRGLPLDLP  
A

25

30

15 SEQ ID No 31 (>ORF13)  
VTIASIDHRDQDLMTGPQAKAPARAAAPDAAPSRRVWAGRVLSGLATLFLTFDAAVKVLKLF  
PAEASTAELGFP AHLVPTLGYLQIACL VAYLIPRTAVLGAILWTGYLGGAIAIHVRVENPLFS  
HTLFP IYVA AFLWAGLWLRDRRVRALTAS PSSQGR

35

40

20 SEQ ID No 32 (>ORF14)  
MTTKNPRKLFVNLSVRDLKRSMEFFSKLGFEFNPQFTDEKAACMVVSEEAYVMLLVESFFKTF  
MKKEICSTSTHTEGLFALSCSSRAEVDDMVKKAVAAGGSHAMPQDHGFMYGWSFYDVGHHW  
EVMWMDPKAIQP

45

50

25 SEQ ID No 33 (>ORF15)  
MTPSERLDATFAALADPTRRAILARLASGEASVTELAKPFAMSQPAISKHLKVLERAGLISRG  
RDAQRRPCRIEAKPLEDASGWL DNYRRFWEGSYERLDDLLEELKERESKGERSKR

55

5

100

SEQ ID No 34 (&gt;ORF16)

10

VAPASAPAAGGRDAAPFLDEAAQWLRGEQAPASRPAGEGPAGRLPGRVLVADDNADMREYALR  
LLVAEGWTVEAVADGRAALERARAHPPDLVLTDMMPRLDGFGLLRALRADDRTRGVAVVMLS  
ARAGEEARVDSLEAGADDFLVKPFSAKELLARVRIHVELARRRREAEGQRQYLNDLFMQAPGP  
5 IAILRGPEHVFEVFNPLYQRLVGGRSLVGEPIRAALPELEGQGIWELLDAVVRTGEPIVGKEL  
15 PVRLDRRGDGTTEEVFFNFVYQPMRDRDGAVEGVFVFAFDVTDQVRARRRVEALVEALKLADQ  
RKDEFLAMLAHELNRNPMASISLSLTLLDDADGDGPASARYREIARRQMGLVRLVDDLDDVSR  
ITRGTVELRLEDVDLAAVVQSAAAVRPAVEARRHDVSLSVGPGDFGMRADATRLQVVTNLL  
TNAKYTPPGGSISVRLTREAAVGAPEAVLRVDTGRGIPAAMLEKVFDFLTQVDQTIDRSTG  
20 GLGLGLTLVRRLLLELHGGSVAAASAGPGQGSEFTVRLPLGPGAAPQPAPSAGPPPPREGPPPA  
10 QRDEPPPPPAQRAEAEAAAADRRRLVVEDAEDVRRVMRAYIEALGHEVTVAVDGLEGVKLL  
ELRPEVAFVDIGLPGIDGYEVARRARAAPGGEALYLVALSGYGGPDDQARSRRAGFDLHLTKP  
25 VVGATLQDVLTAAPT

25

15 9. DNA sequence according to claim 7 selected from the fol-  
lowing

30

(a) open reading frames, and peptide sequences corresponding to  
said open reading frames:

20

35

pEPOcos6\_ORF1 sequences:

(1) nucleotide sequence

40

Seq ID No 35 (>pEPOcos6\_ORF1.seq)

25 GGATCACCTGCGGCGGATCGCCGACCTCGTGCTGGTGTTCGGCTCGCTGGATGAGAAGCCGG  
CGGCGCTACTGATAGAGACGGCGACGCCGGGCTGCGGTGGAGCGGTTGCGGGAGATGCTCG  
GCTTTCGGGCGGCCCCACCTGGCGAAGCTGTCTTCGACGGTTGCGAGGTCCCCGAGGCTCAGC  
45 TGATTGGCCGGCCCGGCTTTGCGCTGATGTATCTGGCCCCCTACGCCCTGGATTTGCGTGGG  
TCAGCGTCGCCTGGGCCTGCCTGGGCATGATCCGCGCTTGCTGGAGACCTGCGCACAGCACA  
30 TCCTCACCCGCCGACCTTCGGCCACCTGCTAGCCGATCACGGCATGATCCAAACCCTGATCA

50

55

5

10

CCAACCTGGGGATTACACCACCAGGCGACGCTGCTCCACACGCTGCAGGCCTGCCGCGCCAGGG  
ATCGCGGCGACGTGACCGCCTCCGAGGCCACCCTCGCCGCCAAATACCTCGCGTCGCGGACGG  
CGGTCCAGGAGACGACCAACGCGGTCCAGATCATGGGCGCGCTGGGCTGCGACGAGGAGGGCG  
CGATCGCCCGCCACTTCCGCGACGCCAAGACGACCGAAATCATCGAAGGCAGCAACCAGATCA  
5 TCGAGGCGCTGCTGGCCAAGAATCGCCCGCGCGGTGCGGACAACATCGCCGCTTCTCG  
15 ATGCGGAAGTCGAGCCCGGTGCGGCGGAGGCGCACCA

(2) peptide sequence

Seq ID No 36 (>pEP0cos6\_ORF1.pep)

20

10 ITCGAIADLVLVFGSLDEKPAALLIETATPGLRVERLREMLGFRAAHLAKLSFDGCEVPEAQL  
IGRPGFALMYLAPYALDFGRVSVAWACLG MIRACLETCAQHILTRRTFGHLLADHGM IQTLIT  
NLGIHHQATLLHTLQACRARDRGDVTASEATLAAXYLASRTAVQETTNVQIMGALGCDEEGA  
25 IARHFRDAKTTEIEGSNQIIEALLAKNIARAGRDNYRRFLDAEVEPGRAGGAP\*

25

15 pEP0cos6\_ORF2 sequences:

(1) nucleotide sequence

Seq ID No 37 (>pEP0cos6\_ORF2.seq)

30

ATGACGAGCGCGGTCCCGACCGTCAAACCAGCCTGCTCGACGACTTCGAGCGCGTCGCCGAC  
20 GTCGATCCAGAGCGGATCGCCGTCCACGCGAGCGAGACGAGCCTGCGCTATGGCGACATGAAT  
35 GCGCGCGCCAACCGCATTGCCCCACGGGCTACGGGCGCGCGGGATCGGGCCCAATCAAATCGTG  
GCGGTGGCGATGGCCCGCACGCCCAGCTGATGATCGTGCTGTACGGCATCCTCAAGGCCGGC  
GCGGCCTACATGCCCATCGCCCGGACGCGCCCGCTGCGCCGCGATCATATGCTGCGCGAG  
40 AGCCAGGCTGCTCTGATGATCGCCGACGAAGAGATCGCGGGACTCGCGGCCCGGTGCTGACG  
25 CCGGCCGACCCGTTCTTCGCGGCCATGCCGACCACAACCCGAGCCGCGTACGACCCGACC  
GACCTGATTTACGTCTCTACACCTCGGGCTCGACCGCCAGCCCAAGGGCGTGCCATGGAG  
CACCGCGCCGTGTGGAATCGCTGACTTGATGCAGGCCCAGTATCCAATCGACACGCAGGAC  
45 GTGATCCTCCAAAAGACGCCGATCGTCTTCGACGTGTCGGTCTGGGAGCTGTTCTGGTGGCCG  
CTGGCCGGCGCCTCGGTGGCCCTGCTGCCGCAATCCATGGAGAAGTTCCCC'TGGGCGATATCG  
30 GCGACGGTGGCGCGGTGCGGGGTGACGGTGATGCATTTCTGTACCATCGATGCTGATGGCCTTC

50

55

5

10

15

20

25

CTTCAGGTGGTGGCGGGCCGGCCCCAGATGGCGGACCAGATGAAGGGCCTGCGCTACGTCTTC  
TGCAGCGGCGAGGCCCTGGCGCCGGCCCCACGTGTACGCCTTTCAGGAGCACATCAACCGAGCG  
GGCAGCATCAGCTTGACCAACCTCTATGGACCCACCGAGGCGGCGGTGACGTCAGCTACTTC  
GACTGCCCCCGGCGCGTCACTCGCGCGGTGCCGATCGGACGAGCGATCACCGGCATCCAG  
5 CTGCTGGTCATGCGCGACGGCGTGCCTCAGCCGCCCGCGTGGAGGTGAGCTCGCCATCGGC  
GGCGTTGGTTTGGCGCGCGGTACATCTCACGGCCAGACCTGACCGCCGACCGGTTCTGTCCG  
15 CATCCAGGCGGCGACGGCCAGCGGCTCTACCGCACCGGCGATCTGGTGCAGGGACGCGGAC  
GGCGAGCTGGTCTTCTGGGGCGCATCGACCATCAGGTGAAAATTCGCGGTCTGCGCATCGAG  
CCCCGGGAAATCGAGGCCAGATCAGCGCCCATCCCGATGTGGCCGACTGCGCGCTGATTATC  
20 GAGCAGGACTCGGAAACCTGCCCLAGCTGACCGCCTACATTGTCTGGCGCGACCGGGCTTG  
10 ACCCGGAAGGCGCTGCTACAGTTCCTGGGCGCGCGGTGCCCGACTACATGCTCCCGAACCGC  
TTCCTGACCTCTACGGAGCTGCCCCGTGACCGCCAACGGTAAGCGCGACTGGCGCGCGCTGCTC  
GGCCCCGCTCGAGACCCTGCCTCTCCCTTTCTCC

15 (2) peptide sequence

Seq ID No 39 (>pEPOcosf\_ORF2.pep)

30

35

40

MTSAVPTRQTSLLDDFERVADVDPERIAVHASETSLRYGDMNARANRIAHGLRARGIGPNQIV  
AVAMARTPELMIVLYGILKAGAAYNPIARDAPPLRRDHMLRESQAALMIADDEEIAGLAARVLT  
PADPFFAAMPDHNPEPRHDPTDLIYVIYTSGSTGQPKGVMESHRAVWNRLTWMQAQYPIDTQD  
20 VILQKTPIVFDVSVWELFWWPLAGASVALLPQSMEKFPWAISATVARCGVTVMHFVPSMLMAF  
LQVVAGRPEMADQMKGLRYVFCSGEALAPAHVSAFQEHINRAGSISLTNLYGPTEAAVDVSYF  
DCPPGASLARVPIGRAITGIQLLVMRDGVPOPPGVEGELAIGGVGLARGYISRPDLTADRFVP  
HPGGDGQRLYRTGDLVRRDADGELVFLGRIDHQVKIRGLRIEPEIEAQISAHPDVADCALII  
40 EQDSETLPKLTAYIVVARPGLTRKALLQFLGARLPDYMLPNRFLTLTELPVTANGKRDWRALL  
25 GPLETLPLPFS

45

50

55

5

103

## pEPOcos6\_ORF3 sequences:

10

## (1) nucleotide sequence

Seq ID No 39 (&gt;pEPOcos6\_ORF3.seq)

15

20

25

30

35

5 ATGTTACACCCGATTCCCACCGACCGTTTCGCCCTGAGCCGACCGCTCTTTCGCGGGTACCTC  
GCGCACGATCCGATCGTGACGGGCGTGCTGGCGGGCCACCATCCAGGCTGGGTCTTGGTGGAC  
CGCGAGCCCCGAGCCGCGCACGGCGCTGCTGTGGGCCCTTTCCGATCGGCTCTTCTGCGTGGGC  
GCAGCTGACACGCTGACCCCGCACGCGCTGGCCGAGCTGTTCCACGACCGACTGATCCCCCAG  
GCCCGTAAGATCGGGCAGCCGTTTTTCCAGGTTTCAGGGCGAGACGGTCGACACCTGGTCCGGAC  
10 CACCTGCATCAGGTGTCGCCGCACCGACAGTCTCCTTCCGCCAGGCATTCCGCTTCGACCGC  
GACCTCTTCGAGCGGCTGCCAACCAAGCCGGAGCTGGCAGAGGCGCGGCTCGTGCCAATCGAC  
GCGCGGCTGCTGGCCGAACAGGCTGATCTGCGCGAGCGGATACTGGCCTCCTGGTCCAGCGAA  
GCTGCCTTCCATGCGCGCGGTTTTCGGCTTCTGCTACCGCGTAGGTGACCAGCTGCCGAGCGTG  
TGCCTGGCATCGCACGTAGGCGGCGGCGCGGCCGAGCTGAGCATCAACACCGAGCTCGAAGCG  
15 CGCAATCGAGGTATGGCAACGCGGCTGTGCCGGCGTTTCATCGCCGAATCGCTGCAGCGCGGC  
CTGACGCCTTGCTGGGGCACCAGACCTTTTCGCTGCCGTCAATCGCGCTGGCCCAAGAAGCTC  
GGTTTCATCCCGACCTTCACCTTCCCCACCTACTGCTTCGCGACCGGCACCGAACAGCCGGAC  
GACAACTTCCTAGGCGAGCTGTACTACAGGAATCGCGCATCGCCGGAAGTGGGACCGATGAG  
CCGCAAGCGGTTTCGGCTGGCGCGGGGTTGGAGCCTGGCCGGCGACACCGAGCGTGCCGCGAGC  
20 TTCGCCGCACGCGCCCTGGCCGAAAGGTGGGCCGGCCACTCGACTCTGGCCACCGATCCGGAT  
TTCGCCCGATTGCGCGCCAGCGCCGCTGGCCCCGCTCAATGTCCCT

## (2) peptide sequence

40

Seq ID No 40 (&gt;pEPOcos6\_ORF3.pep)

45

50

25 MLHPIPTDRFALSRLFRGYLAHDPIVQVLAGDHPGWVLVDREPEPRTALLWAFSDRLFVVG  
AADTLTPHALAELFHDRLIPOARKIQPFQVQGETVDTWSDHLHQVSPHATVSVFRQAFRFDR  
DLFERLPKPELAEARLVPIDARLLAEQADLRERILASWSSEAAFHARGFGFCYRVGDQLPSV  
CLASHVGGGAAEL SINTELEARNRGMATRLCRRFIAESLQRLTPCWGTETFRLPSIALAQKL  
GFIPTFTFPTYCFATGTEQPDNFLGELYRESRIAGSGTDEPQAVRLARGWSLAGDTERAAS  
30 FAARALAEGWAGHSTLATDPDFAPLRASAAWPRLNVP

55



5

104

10

**pEP0cos6\_ORF4 sequences:****(1) nucleotide sequence**

5 Seq ID No 41 (&gt;pEP0cos6\_ORF4.seq)

15

ATGATTTGTCACTCCACCGCTTCATTTTCTCCACGTTCCCAAGGTCGCCGGCACAAGCGTC  
AAGGACGTCTCGGCAAGAGCTATTCCAGGAGGACCAGGTCACGTTCCAGATCGCTCCCAAT  
CCCCACTACCCACCTGAATGGACTGCGCCTTACGAGGAGCACATTATTGCCGCTGAATTGAAG  
AGCCAGTTGGCGCCGAAATTTGGGACGATTACTTCAAGTTCGCCTTCGTGCGCCATCCGCTC  
10 GACTGGGCGGTCTCCAATTACTTCTTCTTCTGCGCGACCGCAAAGGCCATCCGGCCCACGAA  
TTCCTGGAGCGGAAGGGCTTCGCCGGTACCATGGACATGTTTTTCGGAGCGGCCGGCGCCAT  
CCGCTGGTTCGCCGGCATGCGCTTCAGCCAATGGGAGTTCTTGTGCGACAGCGAGGCGCGGACG  
25 CTGGTGGACTTCGTTGGCAAGTACGAGCGGCTCGAGCAGGACTTCGCCGCCGTGTGTATCCGC  
ATCGGGCTGACCCCGCCGACTTGCCGTGCCTCAACCAGACTCGCCACCAATCCTTTACCACT  
15 TACTACGACGAGGCTTTGATGCGCCAAGTCAGCCGCGCGTTAGCTCGCGATTTGAAATTTTT  
GATTATGCC

30

**(2) peptide sequence**

Seq ID No 42 (&gt;pEP0cos6\_ORF4.pep)

35

20 MICHSHRFIFLHVPKVAGTSVKDVLGQELFQEDQVTFQIAPNPHYPPEWTAPYEEHIIAAELK  
SQLAPEIWDDYFKFAFVRHPLDWAVSNYFFFLRDRKGHPAHEFLERKGFAGTMDMFFGAAGRH  
PLVAGMRFSQWEFLCDSEGRITLVDFVGKYERLEQDFAAVCIRIGLTPDLPLCLNQTRHQSFTS  
YYDEALMRQVSRALARDFEIFDYA

40

25 **pEP0cos6\_ORF5 sequences:****(1) nucleotide sequence**

45

Seq ID No 43 (&gt;pEP0cos6\_ORF5.seq)

ATGAAAGTGGACAAGCGGAATGTCGACGACATTCTCGGACTCACTCCGACACAGACAGGCATC  
30 TTGTACCACTACCTGCTGGACCCGAGGCCGACGCCTATTTGAAACAATTGACGCTGCACCTG

50

55

5

10

15

20

25

30

35

40

45

50

55

GAGGGGCCGCTCGACGTAGCGCGCTTCCGCCGCGCTGGGAGCGCGTGGTGGCGGCTCACGAC  
CAGCTGCGCGCCGTGTTTCGCTGGCAAGGGATCGAACACCCGGTGCAGATCATCCTCAAGCAG  
CACGTGCCGGACCTGGAGTTGGCGGAGGTCCCGCGCGACGCCGATCCGGCAGCCTTCCTGGCG  
CAATGGGTGCGGGCCGACCGGGCGCGCAAGTTCGACTTCGAGACGGTGCCTTTTCGCATCGGC  
5 CTCTGCCGACTGATACCAACATCACGTGATGCTGCTCAGCAATCACCATATCCTGATGGAC  
GGTTGGAGTACGGGCCCTGATTCTGCGGGACTTCCTCGCTGCTACGGCGACTCCGAAAAGTGG  
15 CGGCCACGCACCCGAACGCACTTCAAGGCGTTCATCAAGTGGCACCAGAACCGGCCACGCCGG  
GGCGAGGAGCGATTTTGGCGCGACCTGTTGCGCGATGCGCCGACGGCGGCTTTCCCGCCCTG  
GGCGTCGAAGAAGGCACCCGCCACTCGCTTGACTTCGGCGCCCGCAGCCGCGCTCTCGACGAC  
20 CGCTTGACCCAAGGCTTGCGCGACATGGCTCGCGACCTCGACGTCACCCTCGCCGCGATGCTC  
10 CATACCGCTTGGGGCCTTCTACTCCAGCGCTACCAGAACAGCTGCGAAGTGATATTCGGGACC  
ACCGTTTCCGGCCGCAACGTGAGCTCGCCGGCCTCGACGAGGTGGTGGCTTGTTCATCAAC  
AGGATTCCGTTCCGCTTCTCGGCCGCGCGCGACGACGCGCCGTGAGGCCTTCGTTGCGGTA  
25 CAGCGCAATCTGCTGGCGAGAAGCGAGTTCGAAGCCACCCCGCTGGTGGACATCAAGGGCTGG  
15 AGTGGTCTCGGTCCGGGCGCGGAACGTGTTGACACCATCCTGGTTCATCGAGAACTATCCCTTG  
GACCGCGCTATCTTCGAGAGTGATTCCAGCCTGCGGTTGACCGACCACCAAATCTTCGAGCGC  
30 ACCAATTACGGGCTGACCTGACCATCGAGACCTTCAGCCGTTGCACGTGACGTAGCCCAT  
CGCCGTGACCTGCTGGGCGACGCGGCCGCTGAGCGAATGCTAGATCATTTACCGGCCTGCTC  
CAAGCCATGCTGCGCTTCCCTCACCAGCCGTTGCGCGCCTCGAGATGAAAAGCGAACACGAG  
20 GCCCACCGCGTCTGCACCAACTCAACCAAACGCGTCAGCCGCTGCCGTCCCAATCGGCTTTC  
35 CACCAGTTGTTCTTCGAGCAGGCCAGGCCGATGGGGCACGACCGCGCTGTGGTGGCGCGCC  
ACGCGCTGGACCTACGGCCAGCTGCTGGAACGTGCCCTGCGTCTGGCGGGACGGCTGCAGGAA  
GCCGGCTTCGCCCCGAGGCGATGTGCGCCGCGTCAGCCTCGGCCCGGTTCCGGATCTGATTCCC  
40 GGTITGCTGGGCCCGCTGTTTCGCCGCGCGCGCCTACCTGCCGCTCGATCCACCCCTGCCGGCC  
25 CAGCGCTCGCGGTTTCATCCTCGACGATGCCGTTGCCGCTTCCTGATCAGCGACGCGCCACTC  
GCGGGGCCCACGCCGATCCATCCGACCCTGCCGGCGCCAGCCCCGTTGACGTCAATTTTGCC  
TGTCAGGACGGCGCCGCGCAGCCCGCTACCTGATCTACACCTCGGGCTCCACCGGCCAGCCC  
45 AAAGGCGTCTGGGTTAGCCACCGCAACCTGATCAACTTCCTGACGGGCATGAGCGCAATCCTG  
CCGCTCGCGGCCGACGACGTGTTTCTCTCGCTGACTACCGTGTGTTTCGACATTTTCGGGCTC  
30 GAGACGTGGTTCCCGCTCAGCCCGGGCTGCACGATCGTCTTGGGCACGCGCGCCGAGCAGTTG

5

10

15

20

25

30

35

40

45

50

55

GACCCGGCCGCGGCTGCCAAGGCCATCTCCTGCCATGGCGTCACGGTTTACCAGGCGACGCCA  
TCGCGACTCCAAC'TTCAACTGGAGCACCCACATTTGTCCGCGCCATCGGCTCCCTGACGACC  
CTGCTGGTAGGCGGCGAACCCTCCCAGCCGAGCTGCTGCGGCGCGTACGCGAAGTGACCGAT  
GCGCGTATCTTCAACCTCTACGCTCCCACCGAAACCACCATCTGGTCCACAGCCGGGGAGGTC  
5 ACCGCGGCGGACGTCCCGGATATCGGCCGCCCGATCGCAAATACGGCGTTTTCTTCTGGCG  
CGAGACGGCTCGATCCAGCCCGCGGCTGCTGGGCGAGTTGTGCATCGCCGGCGAGGGCGTG  
GCGTTGGGCTACCACCGACGGCCGACCTGAACCGAGAACGGTTTCGCGAGATTCCGCCGGGC  
CGCCTGCCCTTTGCCGGCAAGCTCTACCACACCGGCGACCTGGCCCGCTGGACCGAAGACGGA  
CGGCTCCTCTGCCTGGGCCGTCTGGACGACCAGCTCAAAGTGCGCGGCCATCGCGTCGAGCCG  
10 GCGGAGATCGAGGCAGTGATGGCGCGCCACCCGGCGGTACGCAGGCGGTGGTCGTACCCGG  
CCGCGCAACGGCGAGCCGGTCTTGCTCGGGTTCTGGACTGCGGAAGGTGAGCCGATGCCAGAG  
GAAGCGCTGAGCGCTTACCTGGCCGACCGACTGCCGAGCTACATGGTACCCGAACGGTGCATC  
CTCATGAAGGCCATGCCGCTAACCGGCAACGGCAAGATCGACCGGCGCGCCCTACCCAATCCC  
TTCCGCTTGACCGAGTCGACCCGGCAGGCGGCGCGCACCTTGGCCCGCACCGCCGGCGAG  
15 CATCGGGTTGCCGAGCTGTGGCAGGCCTTGTTGCGACGCGAGGCGATCGGCTTGGACGAACCC  
TTTTTTTAGGCGGCGGGAACCTATTTCGGCTTGATTTCGGCTTACGCCAAGCTGGAATCCGCC  
TTCCGGAAGTCGTTCCCGATCACCGATTGTGTTCCAGCATAACAGTATTTCGACGCCAGGCAGAA  
ATGCTGAGCGGCTCGTCCGTGAGGCGCGCTCGCGGAGCCGTGCCGCAACCCCGGCGGCC  
GCCGCCAAGTTGCCTCCTCGGCAGCTAAATCCCCAGGGGAGCGCGGCGGCGAGCGACGTG  
20 AGCGGCCTGACCGCGCAACCGCCCCAACCCCACTTCCGGCCCATCGCCGTTATCGGCCTCGCC  
GGCCGATTCCCCGCCGACCCGACCTCGACGCCTTCTTGAAGTGTCTACGAGGGTTCGCTGC  
GGCATTCGCTTCTTACGCCAAGCCGAGCTGCGCGACGAGGGTCTCGACGCGAATCGAATCGCG  
TGTCATAACTATGTCCCGGCCAAAGGTTTCTCGACCGGGCCGACCACTTTGATGCCGACTTC  
TTCCGCATCCCGCCGCGGACGCAGAAATCACCGATCCGCAAATTCGGCTTCTGCTTGAGTGC  
25 TGCTGGAACGCGCTGGAGCATGCCGCTACCCGCCCGGCGGCGGAGATCGGGCTCTTCGCC  
GGCTCCTCGGCCAACTATCACTGGCTCGAATACGTGGGCATTTCCGAGGAGAGCAGCAATCGA  
TTCCCGGTGATGATTCAAAACGAAAAGGACTACCTGGCCACGCGGATCGCCTACCAGCTCGAT  
TTGAAGGGCATTGCCGTACCGTGCAAACGGCCTGCTCGTCGCTGACCGCGGTGAGCTG  
40 GCCTGCGATGCGTTACACGCCGGCCGCTGACCATGGCTTTGGCTGGTGGCGTTGGTCTGACC  
TATCCGTTGCGCGCCGATACCTGCACGAGGATGGAATGATCTTCTCCCCGACGGTTCGGTGC  
50  
55

5

107

10

15

20

25

30

35

40

45

50

55

CGGGCCTTCGACGCCCAGGCGGCCGGCACGGTCTGCGGCAACGGTCTGGGCATGGTGGTGCTG  
AAACAGCTCGACGCGGCGCTGGCCGACGGCGATGCCATCCACGCTCTGATTAAGGGCATCGCG  
GCCAACAAACGACGCGCGGCCAAGATCGGCTACACGGCGCCCTCGCAGAACGGTCAGGCGCGG  
GTGATCCGCGCCGCCCATAGGCTCGCCCAAGTCGCGCCGGAGACCATCGGCTATGTAGAAGCC  
CACGGTTCGGGCACGCCGCTGGGCGATCCGATCGAGGTGGCGGGCCTGACCGAGGCCTTTGAC  
AGCCCCGCTCGCGGCTTCTGCGCCTTGGGTTCGGTCAAGTCGAATGTGGGTCAATTCGATGCG  
GCAGCGGGCATCGCGGTTTCATCAAGGCGGTGCTCTCGCTGTCCCATCGGACCCTGTTTCGCC  
AGCCTCCACGTGCACACGCCCCAACCCGAGATCCCGTTCGCCGACGGTCCGTTCAGGTCAAC  
ACGGAGACCCGGCCCTGGCCAGCTGCCGACCATCCCCGCCGCGCCGGCGTCAGTCTTTCGGC  
ATCGGCGGCACCAACGTGCACGCCGTCTTGAAGAGGCGCCGAGTTGGCCGAGCACGCGGGG  
CGGCGGCGCGAGCGGCAGCTGTTCTGGTCTCGGCGCGGACTGCAGCCGATCTGGAGCGACGC  
ACCGCGGCGCTGGTCCGCCACCTGGCCGCGCATCCGGACCTCGCACCAGATGACGTTGCCTTT  
ACCTTGACACGCGGGCCGCAAACCGATGACCCACCGTCGTTTCTGGTCGCCGCCGACCTCGCG  
GAAGCCGCCGCGCTGCGCCGAGCCCGATCCAGTCAAATCCGCCGCGGCGCGCGCCGACCGC  
TGCCAGGTCTGGATGTTCCGCCGTCTCGGCTCTCAATACCCCGCATGTGTGGCGGCCTCTAT  
CGCACCGAGCCGGCCTTTCGCGAGCAAGTCGACCGCTGTTTCGACCTCCTCGCGCCGCGTTGC  
GATTTGAAGCCCTCGCTCTTCCCCGAGCCCGATCAGGCCATCGACGCATCAGCCCTCGCGGCC  
ATCGACACCGCCAGATCGCCGTCTTCGTCTGCGAATACGCGCTCGCACGGATGCTGGAAGGC  
TGGGGGCTGCGTCCGATCGGCTGATCGGTTACAGTTTCGGCGAATACGTGGCCGCCTGCCTG  
GCCGGCGTCTTCTCCCTGCCCCGACGCCTTGGCAATCGTCCGCGAGCGTGGCCGATCCTGGCG  
GCGGCCGAGCCGGGCGCGATGGTCAGCGTGCCCCCTTCCGGCCGAGCGCGTCGCGTCTGCTG  
GAGCCGCCGCTTGCTTGGCCATTGACAACGGCCCCCTCATGCGTGGTGTCCGGGCCGGTCGAA  
CCGGTGCGCACCTTCACCGCTCGCATGAAGCCGGACCGGGTCTGGGTGACGCCGCTCCAGGCC  
GAGCGCCCGATGCATTGCGCGCTGATGGCCGAGGCCGGCGGCTCACTGCGCGCCATGTTGGCC  
GGGTTCCGCCTGAATGCGCCGCGAATCCCGATCTTAAGCAATGTTACAGGAACCTACCTAACC  
GACGAGCAGGCCCCGAGACCCGATTACTGGGCCCGTCACCTGTGCGGCAACGTTTCGCTTCGCC  
GACGGTGTGCGAACCTTGTGGCCGAGCGCGATCCGGTGTTCCTTGAATTCGGGCGCGGCCGC  
GATCTGAGCTCCTTGGTGCGCCACCAGATGCCGAAGGCGCCGACGAGCCGATCGCACTGATC  
CGTCATCGCGAAGATCCGGTGCGCGACGAAGACCTCCTGCTCGATGGCTTGGGCCGCTGCTTC  
CTGCGTGGGGCGACCCTCCACGGGCAGGCCTTGTACGCCGGCCGAGGCTGCCGCCGCGTGCCG

5

108

10

15

20

25

30

35

40

45

50

55

CTGCCCGGTTACCGGTTCCAGGGTCCACGCTGCATGCCGGCCCCGCGCCGGACTGCCCGGCCTG  
GCGCGACCGACCGTGGGAGCGACCAACCATCAGCTACCGACAGCCTGGAAGCGGCGCCGCGC  
TTGGCGGCTGTGAATCGCTCGCGCCGCAATCCTGGTTGGTATTAGCGACGGCAGCGAATTG  
GCGGGCGAGCTGGTGGCCGGCCTGCGCGCTTCCGGTTGCGCGACCAACCTCGTGAAGGTGGG  
5 CTGGCGTTTCGCGCGCTTCGCGGGCGGCTTCCGCGCGAATCCCCGCGAGGAACAAGATCTCGCA  
15 CAGCTGTTTCGCGACCTGTTCGGCCGAAGCGATGCTGCCCCACCCACATCCTGCACCTGCTCAGC  
CTGCCGTCGCCGGAGCGCGACTCGCCGCTGGCGCGCCTGGAGCACCTCACCGAGCTGGGCTTC  
CACCATCTGCTGGCCCTGGCCCGCCAACTGGAGGCGGTGCGCGCCCCGAGGTCCGCCTCGCC  
GTGGTGACAACCGGCCCTGGCGGCGATTGGCGGCGAGTCCGAGCTGCGGCCCCGAGGTGCGGCTG  
20 TTGCGGGGACCTGTCCGCGTGATTCCCTTTGAATTCCCGAACTTGGCGCTGCGCCTGATCGAC  
10 CTCGACTCGGCCGATCCCATCTGGCGTAGCGGTTGTGAGCCGTTGCTGCGCGAAATGGGCGCT  
GCCCCGGGACCTGAAGAAATCGCGCTGCGCGGCACACCGGTTGGGAGTTGGGCTACGAGCCG  
GTGAGGGGGGCACCGTGAGCACCATCTCCTCGCGACTGCGCGAGGGCGGCGTCTATCTGATC  
25 ACCGGTGGCCTCGGCGGCCTGGGTCTGGCCTTGGCCCCGTACCTCGCCCCGAAGTACCGCGCC  
15 ACCCTGATCCTCGCTGGCCGGCGAGGCGCGCCGGCGCGGAGCTCTGGCACAGGCGCCAGCG  
GAGTTCGTACCGGTTCGAGCTGCGATCGCACAGATGGAGGAGTGTGGCGCCCGCGTGATTCCC  
30 GTGCGCTCGACGTACCGACGCGGACCAAGTGAACGCGTTGTTCCGCCACCATAGAAGCTACG  
GTGCGCAAGATTGAAGGCGTTTTCCACATGGCTGGCATCGTTGACGGCGGCATCATTCGAACG  
CGCACGCGCGCTGCCAGCGACGCCGTGCTGGCGCCCCAAAACGGTCGGAACCTGGATTCTCGAT  
20 CGGGCTCTCCGCGGCGCCGGTGGCGCGCTTCTGGTGCTGTACTCCTCGATCAACGCGGTCGTC  
35 GCGCCCTTCGGCCAGGTTGCCTACGCCGCCGCAACGCCTTCTCTGACGCCTTCGCCAGCGCC  
CACGAACACGACGAGCGTCTTTTCCGCGTCAGCATCGGTTGGGACACCTGGCGCGAGGCCGGC  
ATGGCCGTCGATGCCGCCCCGCGCCCGCGCGACCAAGGCCCGCTCGAAGGGCTTAGCGACGAG  
40 CAGGGCTTGGCGCTGCTCGAAAGCGCCTTGGTTCGGTTGCGAACCAGCGACTCCTCGTCTCCATC  
25 AGCGAACTGCGCGCTCGACTAGCCGAGCATCATCGCAACGGCGGCATTCCCCGGTTGCTCGGG  
CCCCGCGCCAACGAGGCGGGTGCAGCTGATTCCGGCGAGGAGGGCGCCACGCAAGACGCGTCG  
CCGGCCCGTCGCGCCCGTCCCAGTCTGGTTCGTGGCCTTCGCGCCGGCCGGCAACGAGCTGGAG  
45 CGCCGATCGTGGCCATCATCGCGCGCTACCTGCGGCTCGGTTCAGGTGGGCGTCGACGACAAC  
TTCAACGATTTGGGCGCCACCTCGCTCGACCTCATCCAGATCGCCCAACGCCTCGGTGCGGAG  
30 TTGGGCCGCGATGTCCCTGTGCTCTCGCTCTACCAACACCGCACCGTACGCGGGCTGAGCCGC

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55

TTCTCGGCGGCGCGCTCCAATCCGCGCGGTCCGCGTCCCGACGGGCGCTGCCGCACCGGGC  
GCCGCCACGCCGGGGTTGCCACCCCGCGCGCCACAACCGTCGCGCCAGCACCTGGAAAAA  
CGCCGTCAATTGAGGAAAAAGGGGGGCCTTCCCATCATGAG

5 (2) peptide sequence

15 Seq ID No 44 (>pEPOcos6\_ORF5.pep)

20  
10  
25  
15  
30  
35  
20  
40  
25  
45  
30

MKVDKRNVDILGLTPTQTGILYHYLLDPQADAYFEQLTLHLEGPLDVARFRRWERVVAHD  
QLRAVFRWQGIIEHPVQIILKQHPDLELAEVPRDADPAFLAQWVAADRARKFDFETVPFRIG  
LCRTDTQHVMLLSNHHILMDGWSTGLILRDFLACYGDSENWRPRTRTHFKAFIKWHQNRPRR  
10 GEERFWRDLRLDAPDGGFRLGVEEGTRHSLDFGARSRALDDRLTQGLRDMARDLDVTLAAML  
HTAWGLLLQRYQNSCEVIFGTTVSGRNVELAGLDEVVGLFINTIPFRFSAAAATTPVEAFRAV  
QRNLLARSEFEATPLVDIKGWSGLGPGAELFDTILVIENYPLDRAIFESDSSLRLTDHQIFER  
25 TNYGLTLTIETFSRLHVTLAHRRDLLGDAAAERMLDHFTGLLQAMLRFPHPQPFARLEMKSEHE  
AHRVLHQLNQTRQPLPSQSAFHQLFFEQAQADGARPALWCGATRWTYQQLLERALRLAGRLQE  
15 AGFARGDVAAVSLGPVVDLI PGLLGPLFAGGAYLPDPTLPAQRSRIFILDDAGCRFLISDAPL  
AGPTPIHPDPAGASPVVDVIFACQDGAAPAYLIYTSGSTGQPKGVMVSHRNLIINFLTGMSSAIL  
30 PVAADDVFLSLTTSVFDIFGLETWFLSRGCTIVLGTAEQLDPAKAAKAI SCHGVTVYQATP  
SRLQLQLEHPTFVRAIGSLTTLVGGEPPLAELLRRVREVTDAIFNLYGPTETTITWSTAGEV  
TAADVDPDGRPIANTGVFLLARDGSIQPPGLVGELCIAGEGVALGYHRRPDLNRERFREIPPG  
20 RLFFACKLYHTGDLARWTEGRLICLGRLLDQLKVRGHRVEPGEIEAVMARHPAVTQAVVTR  
PRNGEPVLVGFWTAEGEPMPEEALSAYLADRLPSYMPERCI LMKAMPLTGNGKIDRRALPNP  
35 FALTESTRQAAPRTLARTAGEHRVAELWQALLRREAIGLDEFFQAGGNSFGLIRLHAKLESA  
FGKSFPITDLFQHTSIRSQAEMLSGSSVEAPLAGAVPQPPAAAAQVASSAAKSPGERGAAATS  
40 SGLTAQPPQPHFRPIAVIGLAGRFPAAPDLDAFLELLTEGRCGIRFFSQAELRDEGLDANRIA  
25 CHNYVPAKGFLDRADHFDADFFGI PPRDAEITDPQIRLLLECCWNALEHAGYPPGGGEIGLFA  
GSSANYHWLEYVGI SEESSNRFVMIQNEKDYLATRIAYQLDLKGI AVTVQTACSSSLTAVEL  
ACDALHAGRVTMALAGGVGLTYPLRAGYLHEDGMI FSPDGRCAFDAQAAGTVCCNGLGMVVL  
45 KQLDAALADGDIAHVIKIGIAANNDCAAKIGYTAPSQNGQARVIRAAHRLAQVAPETIGYVEA  
HGSSTPLGDCPIEVAGLTEAFDSPRRGFCALGSVKS NVGHLDAAAGIAGFIKAVLSLSHRTLFA  
30 SLFVDTPNPQIPFADGPFQVNTETRPWPAADHPRRAGVSSFGIGCTNVHAVLEEAPQLAEHAG

5

110

10

5

15

20

10

25

15

30

20 pEP0cos6\_ORF6 sequences:

35

(1) nucleotide sequence

Seq ID No 45 (>pEP0cos6\_ORF6.seq)

40

25

45

30

50

55

RRRERQLFLVSARTAADLERRTAALVRHLAAHPDLAPDDVAFTLHAGRKPMTHRRFLVAADLA  
EAAARLAEPDPVKSAARADRCQVWMFAGLGSQYPGMCGLYRTEPAFREQVDRCFDLLAPRC  
DLKPSLFPEPDQAIDASALAAIDTAQIAVFVCEYALARMLEGWGLRPDRLIGYSFGEYVAACL  
AGVFSLPDALAIVRERGRILAAEPGAMVSVPLPAERVASLLEPPLALAIDNGPSCVVGSGPVE  
PVRTFTARMKRDRVWVTPLQAERPMHSPLMAEAGGSLRAMLAGFRLNAPRIPILSNVTGTYLT  
DEQARDPDYWARHLCCNVRFADGVRTLLAERDPVFLEFGPGRDLSSSLVRHQMPGADepIALI  
RHREDPVRDEDLDDGLGRCFLRGATLHGQALYAGRCRRVPLPGYPFQGPRCMPARAGLPGL  
ARPTVGATTISYRPAWKRAPRLAIVESLAPQSWLVFSDGSELAGELVAGLRASGCATTLVEGG  
LAFARFAGGFRANPREEQDLAQLFATLSAEAMLPTHILHLLSLSPERDSPLARLEHLTELGF  
HHLLALARQLEAVGAPEVRLAVVTGLAAIGGESELRPEVGLLRGPVRVIPFEFPNLRRLRID  
LDSADPIWRSGCEPLLREMGAPGPPEEIALRGTSRWELGYEPVEGGTVSTISSRLREGGVYLI  
TGGLGGLGLALARHLARKYRATLILAGRRGAPARELWHQAPAEFVPVAAAIAQMEECGARVIP  
VALDVTADQVNALFATIEATVGKIEGVFHMAGIVDGGIIRTRRAASDAVLAPKTVGTWILD  
RALRGAGGRFLVLYSSINAVVAPFGQVAYAAANAFLDAFASAHEHDERLFRVSIQWDTWREAG  
MAVDAARARGDQAPLEGLSDEQGLRLLESALVGCEPRLVLSISELRARLAEHHRNGGI PRLLG  
PRANEAGAADSGEEGATQDASPARRARPDLVFAFAPAGNELERRIVAIIGAYLRLGQVGVDN  
FNDLGATSLDLIQIAQRLGRELGRDVPVVSlyQHRTVRGLSRFLGALQSARSGVPTGAAAPG  
AATPGVATPPRPQPSRQHLEKRRQLRKKGGPSHHE

ATGAGTGAAGTATCCATTGCCCGGCTTGGACATCGCGGTCATCGGCATGGCCTGCCGCTTT  
CCCGGTGCCCGCAACCTCGCCGAGTATTGGGCCAACCTGATCGAAGGCCTCGAAACGCTCAGC  
TTCTTCAGCGAAGAGGAGCTGCGTGAGGCCGGCTGCGATCCGGTCCAACGGCCAGCACAAAC  
TACGTGCGCACCAAGGGCCTGCTCCCTGACGCAGACCGTTTCGACGCCGATTTTTTGGTTAT  
TCCCCGCGCGAAGCCCAGGTGATGGACCCCCAGATCCGCGTCTTCACGAGGTCTGTTGGCAG  
GCGCTGGAGCACGCGGGCTACAACCCGCATCGCCACACCGGCACGATCGGCCTGTTGCGCGGC  
GCCGCGCCCAACGTTTTTTGGGAGTTTTCTCTCTATCGGTCCGATGCCGCCAATTTAGGCAAC

5

10

5

15

20

10

25

15

30

35

40

25

45

30

50

55

TTCACGCTGGGCCTGCACAACAACAAGGACTACCTGAGCTCGCGCATCGCCTACAACCTCAAC  
CTGACAGGGCCCAGCTACACCCTGTTACCGCCTGCTCGACCTCGATGGTCGCCATCCACCAG  
GCCGTCCAGGCGCTGCTCAACGGCGAATGCGACCTGTGCATGGCCGGCTCGGTCTCCATTACG  
CTGCCACTGGTTGCCGGCTACACCTACACGCCGGCATGATCGTCTCGCCCGACGGCCATTGC  
CGCACCTTCGACGCAGGCGCCAATGGCACTGTCTACGGCGACGGGGCCGGCTGGTCTGTTCTC  
AAGCGGGCCGAGGATGCGTTGGCCGACGGCGACCACATATTTGCGCTCATCAAGGGCTCGGCG  
CTCAACAACGATGGCAGTCGCAAGACCGGCTACACCGCGCCAGCGTCAGGGGCAGGTGGAG  
GTGATCCGCGCGGCGATGAACCTGGCGGAGGTGAGCCGAGCGCATCAGCTACGTGGAACC  
CACGGGACGGGCACCACGGTGGGCGATCCGCTGGAGTTCAGGGCGCTAAAGGAGGCCTTCGGA  
GGTGGCTGCAAGGCCTTCTGTGGATTGGGTTGCGTCAAGCCGAACATCGCCCATCTGGACGTG  
ACGTGGGGATCGCGAGCTTCATCAAGCTGGTCTTGGCGCTGGAGCACCGCATCCTACCGCCC  
ACGCTCCACTTCCAACCTGCCCAACCCGAAGATGGATGTGGTCGATAGCCCCCTTCTACATCGTG  
GCTGAGCGCGAACCTTGGCGGAAGATCTGCTGCCGCGTCGGGCCGGTGTGAGCGCGTTCGGT  
CTGGGTGGCACCAACGTCCACATGATTTTGGAGGAGTTTCAGCGCGAACCGGCGCGAACAGC  
GCGCGCACGCGCCACCTGACGGTGTGACGGCGCGGTGCGCCGAAGCCCTGGCGCAGCTGGCG  
GCCAACCTCGCCGAACACCTGCGCGAACACCCGAGTTGGCGCTGGCCGATGTGGCCCATACG  
CTGTGCGACGGCCGAAGCCACATCCATTGCGCGCATCCTGGTGGCGACCGATACGACGGCG  
GCGATCGACGCCTTGATGAACGACCGCGATCCGCGAACGCGTTTCTTCGAAGCGACCGGGCGC  
GGCGAGTCGGTGATCCTGTGTTTTGACGAAACGCCGCCGAGCCGCGAAGCGCCCCGCTACCTC  
TGGGATCACGAGCCGCTTTATCGCGCGCGCGGCGACGTCGTGCTTGGCTGGTGAGGTGCGCGAC  
CCGATCTGGAAGGCTGCTTTACTGCCCTGATCGCCGAGCAGGGCGCGGAGCCGCTTTTGC  
CACCAATACGCGCTGGCCGGATGGCTGCTGGCCATGGGGTTGACCCCGTCGGCGTTGATCGGC  
GTGGGCCAGGGCGAGTGGGTAGCAGCGGCGCTCGCGGAGGTGTTCCCGCCATCGGCCTGCTTG  
CGCTGGATTAGGTTCCGGCGAACGGCTCCCGCAGCCGCGCGATCAACGGATTCCGTTTCTCTCC  
AATTCTCTGGAACTGGATCGTTGGGCGTGAGTTGGCCGACCCGATTACCCAGAAAGCAG  
AAGGGTAAGCGCTGCATGAAGCGCCGTCGGTCCCAACCTCGGTCAGCTGGTGCAGGATGGGGG  
CGATGGAACCGGCTCGGTGAGCTCGTTCGCGCGCTGCTCTCCGCGGAAGCGGAGCGGGGACG  
GTGATCGGCCCCGAGGGCGAGGTTTCATCTCTGTCGACGAGCCGGGCGGGGTGCGCGCCCAG  
TACCTGGGGGCGAGCTCGAGG



## (2) peptide sequence

Seq ID No 46 (&gt;pEP0cos6\_ORF6.pep)

MACRFPGARNLAEYWANLIEGLETLSTFFSEELREAGCDPVQLAQHNYVRTKGLLPDADRFDA  
DFFGYSPREAVMDPQIRVFHEVCWQALEHAGYNPHRHTGTIGLFAGAAPNVFWEFLSYRSDA  
ANLGNFTLGLHNNKDYLSRIAENFNLTGPSYTLFTACSTSMVAIHQAVQALLNGECDLCMAG  
SVSITLPLVAGYTYTPGMIVSPDGHCRTFDAGANGTVYGDGAGVVVLKRAEDALADGDHIFAL  
IKGSALNNDGSRKTGYTAPSVQGGQVEVIRAAMNLAEEVEPEAISYVETHGTGTTVGDPLEFEAL  
KEAFGGGCKAFPCGLGSVKPNIGHLDVTSGLASFILVLALAEHRILPPTLHFQLPNPKMDVVD  
PFYIVAEREPPWRELLPRRAGVSFAFGLGGTNVEMILEEFQREPAANSARTRHLTVLTARSPQA  
LAQLAANLAEHLREHPALADVAHTLLHGRKPHPFARILVATDTTAAIDALMNDRDPTRFF  
EATGRGESVILCFDETPEPRRSARYLWDHEPLYRAAATSCLAGEVADPDLEGCFALIAEQGA  
AAAFCHQYALAGWLLAMGLTPSALIGVGQGEWVAALAEVFPFSACLWIRFGERLPQPRDQR  
IPFLSNFSGNWIVGRELADPDYPRKQKGRCKMRRRSQPRSAGAGWGRWNRLGQLVARCSSAG  
SGGGTVIGPRARFISSTSRARVRAQYLGAASSR

## pEP0cos6\_ORF7 sequences:

## (1) nucleotide sequence

Seq ID No 47 (&gt;pEP0cos6\_ORF7.seq)

ATGGAACCGGCTCGGTCTAGCTCGTCGCGCTGCTCTTCCGCGGGAAGCGGAGGCGGGACGGT  
GATCGGCCCGAGGGCGAGGTTTCATCTCGTCGTCGACGAGCCGGGCGCGGTGCGCGCCAGTA  
CCTGGGGCGAGCTCGAGGTAGCGGTCCCGCGCCAGTAGGGCATCGCGGAATGACGTCCGC  
CAGGTAGGCCTCCGGGTCGAGCCCGTGCAGCTTGCGCTCGCCACGAGCGAGAAGAGGTTGGC  
CGCGGCGGAGGCGTGGTCGTCGCTGCCGAAGAAGAGCCAGGACTTTCTCGCAACCGCAATGGA  
TCGAGCGCTCGCTCGCTGGCGTTGTTCTCCAGGCGCAGCCGACCGTCGTCGAGGAAGCGCCG  
CAACGGCTGCTCTTGTTGAGGGCGTAGCCGAGCGCGGTGGAGACCAGGCCGCGCTCGCGGGG  
ACGAGCGTGCTCGGCCCTGGCCCAGGCAAAGAACGCGTCGACAGAGGGCGGACGACGACATC  
GCGACGCACCTTGCGCTGCGCGGCGCGCAGGTCCGCCAGCGCGCATCGGCGGCAAACAGGGC  
GTTGATGCGCCGAGCCCTCGACACCGAGCTCGTGCTTGACAGACCGCCGCTCCAGAAAGTT  
GGTACGGCAATGCGACCAGCATCCGACTTCGGTCGGGGCGGACCGCGCTTCTCGTCGGCAGC

5

10

15

20

25

30

35

40

45

50

55

AGCGCCTCTTGGTGGTGTGCCGCGGAAGAGGGCGTCATAGATGGCGTGAGCGTCAGCTTGAAT  
ATACCGAGAGAAGCCGCGGAACATCTCGCAGACCGGGCGCTGGTATGCTTGGGCTGGTACTC  
GAAGAAGACGTGATCCTTGTCCGCGAGGACGACGAAGAAGTGTCCCTTGCGGCACGGCCCCGGG  
CTTCTTGTCTTGGCTCCTGGATGGGCCCAGGCTGGACGGAGACCCCGGTGGCGTCCGTGGA  
5 CAGGCAGAAGCGGTCTCGAAGGCCTCTTTGCGCGCGGCCTCGACGATGGCGCCCAGGGTCGC  
ACCGACGTCTTCGGCGTAGCGGCACATCGTGCCGCGATCGAGCGACGCGCCCTGAAGCTCCAG  
15 CTGCTGCTCCAGTCGATAGAACGGGACGCCGAGCAGGTAAGTCTGCTGGTGAGGATGTGCGCAAT  
CATCGACGGCGCGAGGAACGACCCCGGAACAACTCCTTCGGAAGCGGCGTCGTGATGAAGAC  
CGTGACAGTCTCGCCCTTCGGCGCGCGCGCGCGCTCGAGCGAGGGCGCGTCGAGCGCTGT  
20 GGAGGAAGCGCTCGGCTCGCCGGCGCTGCCGTGTCTCCGGCCTGACGCTCGCAGCGGGCGT  
CGGCGTCGGGGCTTCTCTCGCGACGACCTGGAGCGGGCGCGCTTCTCCTCGCCCCGAATGCT  
CGCATCCGTGACGGACCGCTCGGCCTTGTACACGACGCGTGCGAGCACGATGCGGCGCATTCC  
25 GCGCGCTCGTAGCCGAGTCGCGAGGTCTCCTCGACCCCGATGCGCGTCGCCGTGCGATCGAG  
CTCGGGGCAGGAGAGCTCGATGCGGACGACGGGCAGGTGCGACTCGGACAGGTGCGGACGGCC  
15 CTTGCCGCCGGACCTTCGTTTCGGCCCCCTTGGGGTCGTGCTGCTGCCGCTCGTCGCCTGTATT  
GCGCTCGGCGGCGTCGAGTGCCCTTCGCGAGGCGCTGGACCTCGAGGAACATCGAGTCGAACGC  
30 CAGCTGCTCCGCGCTCACCTCGGCGCGCTCCGCTTGGCCACGAACAGTCGACGTGCGAGAAG  
CTGCAGCTGCTCGAGCGCACGGGTGTAGGCGCGCCGAAGCTGCGCGAGCGCATCGCGCGCTCC  
CACGAGCTCGCTCTTTGCCGCGGCGAGCTCCGCTTCGAGCTGCGCGATGCGCTGCTGCTCGGC  
20 CGAGAGCGTCGGCTTGGCGGCGGCGTCTGTCACGACGCCGCTCTACGTAAGCCGCGCGTACTT  
GTGAGCGAATTTCGTGCGGCTCAGTGGACGCGGCGCGGTGCGCGCCTTCGCGGTTTGGACGTG  
GGCGGATCTCGATGCCGTGAGCAGCGTCTCGAGCGTGGCGTCTCCACCTCGACGTGCGTG  
GCGCCCTCGGTGCGGGGGTCCGGAAGTGCGAACGCTCCGCGATCAAGGCGTTTTGAAAACAGG  
40 CAGATTCCACTGCCATCGAAGAAGAGAACTTGATCGTGGTCCGCCGCTTGCCGACGAACGCG  
25 AACAGCGCTCCGCGAGCGAGCTCGTACCCACACGCTCACGGATGAGACCCGAAAGCCGCTCG  
AAGCCG

## (2) peptide sequence

Seq ID No 48 (&gt;pEPOcos6\_ORF7.pep)

MEPARSARRALLFRGKRRRDGDRPEGEVHLVVDEPGAGARVPVPGGELEVAVPRPVGHRANDVG  
QVGLRVEPVQLAARHEREEVGRGGGVVAAEEEPGLSRNRNGSQRS LAGVVLQAQPTVVEEAP  
5 QRLLLVEGVAERGGDQAALAGTSVLGPGPGKERVDQRADDDIATHLALRGRQVRQRAIGGKEG  
15 VDA PQPLDTELVLADRR LPEVGTAMR PASDFGRGRTALLVGSSASWWCAAEECVIDGVS VSLN  
IPREAAEHLADRGAGMLGLVLEEDVILVREDDEEVSLAARPGLLV LALLDGPRLDGD PGGVVRG  
QAEGLLEGLFARGLDDGAQGR TDVFGVAAHRAAIERRALKLQ LLLQSIERDAEQVLAGEDVRN  
HRRREERPPEQLLRKRRRDEDRAGLALRRRRRRRVERGRVERCGGSARLAGRCRVLRADARSGR  
20 RRRGFSRDDLERGRFLLARTARIRDG PLGLVHDACEHDAHSAALVAESRGLLDPDARRRRRIE  
10 LGAGELDADDGQVGLGQVATALAAGPSFRPLGVVVLPLVACIALGGVECLREALDLEEHRVER  
QLLRAHLGALRLGHEQSTSQKLQ LLERTGVGAPKLRERIARSHELALCRGELRFELRDALLG  
25 RERRLGGGVVHDAALRKPRVLVERIRAAQWTRRGARLRGLDVGAISMPSSSVSSVASSTSTCV  
APSVGGSGSANAPRSRRFENRQIPLPSKKRILIVVRLPTNANSAPQRASYPTRSRMRPESRS  
15 KP

## pEPOcos6\_ORF7.1 sequences:

## (1) nucleotide sequence

Seq ID No 49 (&gt;pEPOcos6\_ORF7.1.seq)

ATGTTCTCTCGAGGTCCAGCGCCTCGCGAAGGCACTCGACGCCGCCGAGCGCAATACAGGCGAC  
35 GAGCGGCAGCACGACGACCCCAAGGGGCCGAAACGAAGGTCCGGCGGCAAGGGCCGTCGCGAC  
CTGTCCGAGTCCGACCTGCCCGTCGTCCGCATCGAGCTCTCCTGCCCGAGCTCGATGCGACG  
40 GCGACGCGCATCGGGGTCGAGGAGACCTCGCGACTCGGCTACGAGCGCGGCGGAATGCGCCGC  
25 ATCGTGCTCGCACGCGTCGTGTACAAGGCCGAGCGGTCCGTCACGGATGCGAGCAGTTCCGGGC  
GAGGAGGAAGCGGCCCGCTCCAGGTCGTGCGAGAGAAGCCCCGACGCCGACGCCCGCTGCG  
AGCGTCAGCCCGGAGGACACGGCAGCGGCCGGCGAGCCGAGCGCTTCCTCCACAGCGCTCGAC  
45 GCGCCCTCGCTCGACGCGCCGCCGCCGGCGCCGAAGGGCGAGACCTGCACGGTCTTCATCACG  
ACGCCGCTTCGAAGGAGTTGTTCCGGCGGTCTTCCTCGCGCGCTCGATGATTGCGCACATC  
30 CTCACCAGCAAGTACCTGCTCGGCGTCCCGTTCTATCGACTGGAGCAGCAGCTGGAGCTTCAG

5

115

10

15

20

25

30

GGCGCGTCGCTCGATCGCGGCACGATGTGCCGCTACGCCGAAGACGTCGGTGCGACCCCTGGGC  
GCCATCGTCGAGGCCGCGCGCAAAGAGGCCTTCGAGACCGCCTTCTGCCTGTCCACGGACGCC  
ACCGGGGTCTCCGTCCAGCCTGGCCCCATCCAGGAGCGCAAGGACAAGAAGCCCGGGCCGTGC  
CGCAAGGGACACTTCTTCGTCTCTCGCGGACAAGGATCACGTCTTCTCGAGTACCAGCCC  
5 AAGCATACCAGCGCCGCGGTCTGCGAGATGTTCCGCGGCTTCTCTCGGTATATTCAAGCTGAC  
15 GCTCACGCCATCTATGACGCCCTCTTCGCGGCACACCACCAAGAGGCGCTGCTGCCGACGAG  
AAGCGCGGTCCGCCCCGACCGAAGTCGGATGCTGGTTCGATTGCCGTACCAACTTCTGGGAG  
GCGGCGGTCTGCAAGCACGAGCTCGGTGTCGAGGGGCTGCGGCGCATCAACGCCCTCTTTGCC  
GCCGATCGCGCGCTGGCGGACCTGCCGCCCGCGCAGCGCAAGGTGCGTCGCGATGTCGTCTC  
20 CGCCCTCTGGTCGACGCGTCTTTGCTTGGGCCAGGGCCGAGCAGCTCGTCCCCGCGAGCGC  
10 GGCCTGGTCTCCACCGCGCTCGGCTACGCCCTCAACCAAGAGCAGCCGTTGCGGCGCTTCCTC  
GACGACGGTCGGCTGCGCCTGGAGAACAACGCCAGCGAGCGAGCGCTGCGATCCATTGCGGTT  
GCGAGAAAGTCTTGGCTCTTCTTCGGCAGCGACGACCACGCCTCCGCCGCGGCCAACCTCTTC  
25 TCGCTCGTGGCGAGCTGCAAGCTGCACGGGCTCGACCCGAGGCCTACCTGGCCGACGTCATT  
15 CGCGCGATGCCCTACTGGCCGCGGGACCGCTACCTCGAGCTCGCCCCAGGTACTGGGCGCGC  
ACCGCGCGCCGGCTCGTCGACGACGAGATGAACCTCGCCCTCGGGCCGATCACCGTCCCGCCT  
CCGCTTCCCGCGGAAGAGCAGCGCGCGACGAGC

(2) peptide sequence

35

40

45

50

55

20 Seq ID No 50 (>PEPocos6\_ORF7.1.pep)  
MFLEVQRLAKALDAAERNTGDERQHDDPKGPKRRSGGKRRDLSES DLPVVRIELSCPELDAT  
ATRIGVEETSRLGYERGGMRRIVLARVVYKAERSVTDASSSGEEEAAPLQVVAREAPTPTPAA  
SVSPEDTAAAGEPSASSTALDAPSLDAPPPAPKGETCTVFITTPLPKELFRRSFLAPSMIAHI  
40 LTSKYLLGVPPFYRLEQQLELQGASLDRTMCRYAEDVGATLGAIVEAARKEAFETAFCLSTDA  
25 TGVSVQPGPIQERKDKKPGPCRKHFFVVLADKDHVFFEQPKHTSAAVCEMFRGFSRYIQAD  
AHAIYDALFRGTPPRGAAADEKRGPPPTVGCWSHCRTNFWAAVCKHELGVGLRRINALFA  
ADRALADLPPAQRKVRDVVRPLVDAPFAWARAEHARPRERGLVSTALGYALNQEQLRRFL  
45 DDGRLRLENNASERALRSIAVARKSWLFFGSDHASAAANLFSLVASCKLHGLDPEAYLADVI  
RAMPYWPRDRYLELAPRYWARTRARLVDDMNALGPITVPPPLPAEEQRATS  
30

5

116

**pEPOcos6\_ORF7.2 sequences:**

10

**(1) nucleotide sequence**

Seq ID No 51 (&gt;pEPOcos6\_ORF7.2.seq)

15

5 ATGATTCCGGCGGGCGTGCAGGTGTTTCGTCGCGCTGGAGCCGGTGGACATGCGCTACGGCTTC  
GAGCGGCTTTCCGGTCTCATCCGTGAGCGTGTGGGTACGAGGCTCGCTGCGGAGCGCTGTTT  
CGGTTTCGTCGGCAAGCGCGGACCACGATCAAGATTCTTTCTTCGATGGCAGTGGAAATCTGC  
CTGTTTTCAAACGCCCTTGATCGCGGAGCGTTCGCACCTCCCGACCCCCGACCGAGGGCGCC  
ACGCACGTCGAGGTGGACGACGCCACGCTCGAGACGCTGCTCGACGGCATCGAGATCGCGCCC  
10 ACGTCCAAACCGCGAAGGCGCGCACCGCGCCGCGTCCAC

20

**(2) peptide sequence**

Seq ID No 52 (&gt;pEPOcos6\_ORF7.2.pep)

25

MIPAGVQVFVALEPVDMDRYGFERLSGLIRERVGYEARCGALFAFVGKRRTTIKILFFDGSIG  
15 LFSKRLDRGAFALPDPPTEGATHVEVDDATLETLLDGIEIAPTSKPRRRAPRRVH

30

**pEPOcos6\_ORF7.3 sequences:****(1) nucleotide sequence**

20 Seq ID No 53 (&gt;pEPOcos6\_ORF7.3.seq)

35

ATGACAAGGACGAAGGCGACCGAAGTGATGTGGTCCGAGCGCGTTCGGGCGTGGCGCGAGAGT  
GGTGAAACGGCGGAGGAGTTCGCTCGGAGCCGCGATTTCGCGCCTCGACGCTGCACGGCTGG  
TCGAGCCGGCTGTTCGCGGGCCGAGCCACCGCGCTTCTGCGCCTGGTGCCGAAGGCGCCCGCC  
40 GTCACGAGCAGCGCTGCGGAGCTCGTCGTCGAGGTCGGCGGCGCGGGGTGCGCGTCGCCGCG  
25 GGTTTCGACCCCGCGCTGCTGGCGGAGGTGGTCCGTGCCCTCGGCGGAGCGGGGCCA

40

**(2) peptide sequence**

Seq ID No 54 (&gt;pEPOcos6\_ORF7.3.pep)

45

MTRTKATEVMWSERVRAWRESGETABEFARSRGFAASTLHWSSRLSRAEPPRFLRLVPKAPA  
30 VTSSAAELVVEVGARVRVAAGFDPALLAEVVRLGGAGR

50

55

5

10

## pEP0cos6\_ORF8 sequences:

## (1) nucleotide sequence

5 Seq ID No 55 (&gt;pEP0cos6\_ORF8.seq)

15

20

25

30

ACTGGACAGCGCAGCCGGGGTGAGACGGCGCTTCGCGCAGCGCTTACGCAGAAGGCGCGCCGC  
GCGCCATTGTCGGATGCGGTGCGCGACTTCGCCGCCGATCGGCTGTTGCTGGAAGTGGGACAA  
CCACTGGACGTAACGGCTGAAGCGAGCCAACGGCTCCAGCTCCGCGGGGCGACCTGTTCCGGC  
GCCTACCAAGCGTTGGCCCAGCTCTGGATCTGCGGCGCCCTGGCCGAACCGCCGCGACTGTAT  
10 CCGACGAACACCGCCGGCGCGTGCCTGCGGAGCTACCCCTTCGAGGGAAAGCGTTCTGG  
ATCGAGGGCTCGCCGTTTCAAACCGCGCCCGCCGCGCCTACCCCAACCGCCGATTCTG  
GGGGACATTCTCAAGGGCGACCCGGCGGACTGGTACTATCGGCCGCGTTTCAAAGCGGCGCCG  
CTCTTGCCAGCCCGTTTCGAGAGCGAACCCGGCGATTGGCTGGTGTTCGAAGATGAGCTGGGG  
25 CTGCGCGCCTGGCTGAGCGAGACCTTGCGCGACAAGGGCGCGCGGTTCGCGACAGTCGTTCTGA  
GGCACCGAGTTCCGACGCTTGGCGTCACAGCGCTTCCAGCTTCGTCCCGATCGACGGGACGAT  
15 TACCGGACCCTGCTGCACGAGTTGAAGGCGCAGGGCATCGCGCCGTTCCACCTGTGCCACCTA  
TGGAGCCTGACCGCCGACCGGATGCCGAGCAGTTGCTCGACGTCAGCTTTCACAGCCTGGTC  
30 CATTTGGCGGCCGCTTTGGGTTTCGGTTGGCTACTTCCACGCCATG

## 20 (2) peptide sequence

35

Seq ID No 56 (&gt;pEP0cos6\_ORF8.pep)

40

TGQRSRGETALRAALTQKARRAPLSDAVRDFAADRLLLELGQPLDVTAEASQRLQLARGDLFG  
AYQALALWICGALAEPPRLYPDEHRRRVPLPSYPFEGKRFWIEGSPFETAPAAGASPPADS  
40 GDILKGD PADWYYRPRFEAAPLLPSPFSEPGDWLVFEDELGLGAWLSETLRDKGARVATVVR  
25 GTEFRRLASQRFQLRPDRRDDYRTLLHELKAQGIAPVHLCHLWSVTAAPDAEQLLDVSFHSLV  
HLAAALGSGVGYFHAM

45

50

55

5

118

## pEPOcos6\_ORF9 sequences:

10

## (1) nucleotide sequence

Seq ID No 57 (&gt;pEPOcos6\_ORF9.seq)

15

20

25

30

35

40

45

50

55

5 ATGAAGTTGAACGTGGTCGCCAACCGGCTATTGACCCCGAGTCGCCCCGAGCGCACCGAGCCC  
GCCAAGAGTCTGTTGCTCGCGGTGACCAAAGTCTGCGCGAAGAGGTGCCCAACGTTTGAACC  
CGCGCCATCAGCGTGGACCTGGATCGCTCGTTGACGCGCGCGCGCCCGCTGGGCCGCCAGT  
TTGTTGGTTGAATGCGGCGCGCCCGTCCAGGAAACGGTGGTGACCTACCATGGCGCAGCCCCGA  
TGGCTGCGCCGCTTCGATCGCGTTGCGGTGAATGGTCTCGGCCCGTTCCACCCCGATCAACCT  
10 GCGCCGCTGCTGCGCGAGCGCGGCGTGTACCTGATCACCGCGGCGCTGGGCGGCGTGGCTGGC  
CAGTTGGC3CGCTACCTGGCGCGGCGCTGCCGGGCGCGGTTGGTGCTACCGCGCGCCCGCCCC  
CTGCCCCGAGCGCGACCAAGTGGGATCGGGAGTCGGCCGTGCTGTCTATGGGACGACAAGACGCGC  
CAGCGCATCGAGCTGGTGCGCGAGCTGGAGCGGCTGGGGGCCGAAGTATTGGTGGTGGCTGCC  
GATGTGCGCGACGAAGCGGCCATGGCGCAGGCGATCGAGGCCTCACTGGCGCGATTTCGACGCT  
15 TTGGACGGCTTGATCCACGGCGCCGGGATCGTGCGGGTCGCGTCGGGCCGACGCCGATCGGG  
AGTATGACGCGGGCCATGTGCGAGGAGCAGCTCCGCCCAAGATGTGGGCCTCGACGTCGTC  
GACCGCCTCTGCGCGATCGCCGTTGGAATTCGCGATTGCCATCTCGTCGCTCGCCCCGATT  
CTCGCGCGCCTCGGCCACGTGCGCTACGCGCGCGCCAACTCTACATGGACGCGTTTCGCGACG  
CGCGCCGCCCGCCGCAACGCGCCTTGATCGCGCTGAACCTGGCCGAGTGGGAATACGAGGGC  
20 CCGGCTACCTACGACGAGCGGGTGGGCCGTTTCGCTCAAGCAGCTCGAGCTCACCAACGAGGAG  
GGTATCCGCGTCTTCCAGACGGTGTGGCCTTGCGCGCGCGCGCCCGCTACAGCAGATCATT  
ATTTCCACCGGCGACCTCCAGGCCCGCTCGACAAATGATTACATCAAATCCCTGCATCGC  
CGACCGGGGCCGTTCCAGCTCAGTCGCGGACCGCGGCACCCGAGGGCGGTTTCGGCTCGGAG  
CGCGCCGCCTTCGAGGCCGCTTCGCTGACGCCTGGTGCGACTTCTTCGGGGTTGAAGAGGTC  
40 GACCCGAACAAAACTTCTTCGATCTGGGCGCCAGCTCGCTCGACTTCATCCACCTCGTCAGT  
CGCTTCAGCAAGGCCATCGAACAGCATGTACCGCTCGAGGCCCTGCTCGAACACTCCACCCTG  
CACGACCTCGCCGCCACCTCGCGGGCGACGCGAACACCGACGCCAGCGACGAAGCGCGCATT  
CGCCAACGGCTGCAAGGCGCCAAGTCCGGCGACATCGCCATCATCGGCATGGCCGGCCGCTTC  
45 CCGCTCGCGCCCGACCTGGACACCTATTGGCGCAACCTGGTTCGAGGCATCGACGCGGTTCAGC  
TTCTTCAGCGCCGAGGAGTTGCGTCTGCTGGCGTCAACCGGCCGAGATCCACCACACCAAC  
50

5

10

15

20

25

30

35

40

45

50

55

TACGTGCCGGCCAAGGGGCGCTGC3CCGACCAGGACTTGTTTCGATGCGGCCTTCTTCGAATAC  
ACTGCCAGCGACGCCGAGCTGATGGACCCGCAAAATCGCGTGTTACACGAGGTCGTGTGGCAC  
GCGCTGGAAGACGCCTGTTTCGACTTCAACGGCGATCACGGCCAGGTCGGCCTGTTTCGCGGGC  
GCCTCGCCGAACCTGTGGTGGCAGTTTCGTGGCCAGCTTTTCCGAGGCCGCAAGACGCAGGGC  
5 ATGTTTACCACCACCCTGCTCAACGACAAGGACTCGATCGCGACCCAGATTTCATACAAGCTC  
GGTCTAAAGGGCCCCGCGGTACCTTGTTTACCAGGCTGTTCCACCTCGCTGGTAGCCGTTGAC  
15 GCCGCTGCCGCTCGATCTGGTCCGCTCAATCGGACATGGCCGTGGCCGGCGCGGTCTCGCTG  
ACTCTCCCCGATAAGGCCGGCTACATCTACGAAAAGGGCATGCTCTTCTCGGCCGACGGCCAT  
TGCCGGGCTTTTCGACGCCAACGCCACCGGCATGGTCTTTCGGCGACGGCGCCGGCGCGATCGTG  
20 CTCAAGCCGTTGGACGCGGCCCTGCGCGACGGCGACCCGATCCATGCGGTGATCAAGGGCTGC  
GCCACCAACAACGACGGCGACCCGCAAGCCGGCTACACGAGCGTCAGCGCCCAAGGCCAGGCC  
GAGGTGATCCGCTCGGCCAGATCCTGGCCGACGTGGCGCCCGAATCCATCAGCTACGTGGAA  
25 GCCCACGGTACCGGCACCAAGTTGGGCGACTCGATCGAGATCAAGGCGTTGAAGCAAGCCTTC  
GCCAGCGACAAGAACGGATTTTTCGGCATCGGGTTCGGTCAAGACCAACCTCGGTACCTGATG  
15 GCGGCGGCGGGGATGGCCGGCCTGATCAAGACGTTCTGGCGATGAAGCACCGCCAATTGCCG  
CCATCGCTGCACTGCGACGAAGTGAACCCCGACCTGGAGTTGGAGCGCAGTCCGTTCTACATC  
30 AACACCCGCTGCGCGACTGGGTTGCACCGGGCGGGCCGCTGCGGGCCGGCGTGAGTTTCGTTT  
GGGATCGGCGGAACCAACGCTCACGTTCCTGGAGGAGCCCGACGCGCGAGAGCGGCACG  
CGCATGCGCCACTGGAAATTATTGATGCTGTGCGCGGCCAGCGAGGCGGCGCTCGACCGCCAG  
20 GCGGATAACCTGGCCGACTACCTGGAGCGCCATCCCGAGGCCACCTCAGCGACGTGGCCTAT  
35 TCCCTCCAGACCGGCCGGCGCGTTCTGGCCTGGCGGCGCACGGTCTATGCGAGTACCGCGAG  
GACGCGGTGACCAGTCTGCGCGAGCGACAGGCCAAGCGGTCCAGACAAGTCGCGTCCGCTGG  
GACCACAAGGACGTGGTCTTCATGTTTCCCGGTGAGGGCGCCAGTACCTCAACATGGGCGCG  
40 GACTTATACGTTCATGGAGCCGGTCTTCCGCGAGGTTCATGGACCGCTGCTTCGAGTTGCTGGCC  
25 CCTTTGTGGTCCGAGCATCCGCGCCAGATCCTTTATCCGAGGGCGGGGTGTCGACCCTGCTC  
CACCGGACTGATTACACCCAGCCGATCGTGTTCTGCTTCGAGTACGCCCTCGCCCATTTGCTG  
CTCTCCTGGGGATTGAAGCCGGCCGCGACCATCGGCTACAGCTTCGGCGAGTACGTTTCTGCC  
45 TGCCCTCGCCGGCGTCTTCTCCCTGGAAGATGCGATCCGTCTGGTGACCGAGCGCGGTTCGGCTG  
ATGGCGGCTTTGCCCCGGGGCGCCATGCTCAGCGTCCCGGTTCCCGAATGCGAGCTGCTGCGG  
30 CTGCTGGACGGCTTCCACGCCCAATCGGCGGCCCATCTGGCGCTGGCCGTGACAAATGGCGCC



5

10

15

20

25

30

35

40

45

50

55

TCCTGCATTGTGGCCGGCGAGCAGGCCGCCATCTCGGCCTTCGAATCGATGCTTCGCAAGAAG  
CGTCTGTGACCATGCGGGTCGCGGTACGCCACGCCGCTCATTTCGAGGTATGACCGGCGCG  
ACCGACGCCCTGCGCAGCATCCTGCGGAAGATCCCCCTCTCCGCGCCGACAATTCCCTTCATT  
TCCTGCGTACCCGGCACCTGGATCACTGCACAGCAGGCTACGGATCGCGAGTATTGGGTGAAC  
CACATGTGCGGGACGGTGCAGTTTCGCGGCGGGTCTGACCGAGCTGGGTCAAAACCGCGAGGCG  
GTGTTCTGGAAGTAGGTCCGGGCCGCGACTTGACGTTGCTGGCCACCGCATCCTGGCCGAC  
AGCGCGGCCGTGTTTCGAGCTGGTCAAGGCGCCCGACGGCGGCGACGACGATGGGTTCCTCCTG  
CTGGATCGATTGGCCAAGCTCTGGAGGCTGGGGATTTTCGATTGACTGGGCCGCTTCTACGCG  
GATGAGCGGCGGCGGAACTCTCGCTGCCGGGATATCCGTTTCGAGCGGCGGCGCTTCTGGATC  
GAGGGCAACCCGCTGGAGATCGCCGCCGGCAGGCCCAATGTCCAGGGCCGCTGGTCAAGGCG  
TCGGACATCGGCGCTTGGTTCTACGTGCCGCAATGGCGGCGGTTCGGTGCTCGCCGAGCCGGGT  
ACAACGGCGGCGGGCGCCGCCGTACGGCGGAGCAGGCACGCGTCGTGACCGAGCTACGGCGG  
GGATGCGCGTCCGGCCGCTTGGGCGAGCGGGCCTGCGGACTGAATGGCGGTGCCCCGTCCGAG  
CGTCCGAAAGAAAGTGTAGCGCCAGCCGGTTCGACCAGCGCAGCGGCGCAGACCGGCGCGGAC  
TGCCCCGACACCGACTGGGGAGCCAGCGGCTGTGCCAAAGGACCGGGCCGAGCCGCGGCCGACC  
TGGCTTATTTTCGCCGACGCCGGCGGATTGGCCGAATCTTTCGCCAAGCGGGTTCAGGCCCGC  
GGCGAGAAGCTTTACCTGGTGGCTTCCGGCTCGCGCTTCGAGCGCCTGGCCGAGACCCGCTTC  
CGCCTCGATCCCGGGGCCAAGTCCGATCACCGCTGCTTTCAAGGCGCTCGACGAGGCCGAC  
ATCTTGCCGACCCACCTCCTCGACTTCCGCTCGCTTGACTGCGGCGGGCCCGACGCCGACCCC  
ATGGACCAGGCCGGCTTCTTCGGGCTGTTGCACCTGGTCCAGGCGATGGCAGAGGCCGGCTAC  
AGCCATCCCATTCGGCTGCTGATCGTCAGTTGCGGCGTCTACGATGTCACCGGTGCCGAACCG  
CTGCAGCCGGCGCGGGCCACGATGATCGGACCGGCTCTGTGCATCCCGCAACAGTATCCGCAC  
CTCGAAACGAGCCATGTGGATTTGGGCGTGGTCCATGCCGACGAGCTCCACGCCGCGCGCCAG  
CTCGACAGCCTACTTGCCGAATGCCTAAGTGCAACGGCCGAGCGCCAATTGGCGCTGCGCGGC  
CGACACCGCTGGCTGCTGGACTACGAGCCAGTCCGCTTGCCGCCGCTCGACCCGGGCCGCTG  
CCCTGGCGCCAGCGCGGGTCTACTTGATCACCGCGGTTTGGGCGGGATCGGCCGCATCCTG  
GCCGAACACCTGGCCCCGACGACCTCGGCTCGCCTGGTCTTAATCGGCCGCGAAACCTGCCC  
GACCGGACGACTGGGACGCTGGCTGAACCGCCCGCAACCGGTTCGACGCCACCCACGAACGG  
CTGCTGCACAAGATCCGCGGATTCGCGATCTGGAAGCGCTAGGCGCCGAAGTCTGTGCTC  
GCCCGGACGTCGCCAACGAAGCCGCCATGCGCGAGGCTACGATCGCGCCGAATCCCACTTC

5

10

15

20

25

30

35

40

45

50

55

GGCACAAATCCACGGGGTGATTACAGCGCCGGCCTGATGGACGCGCAAAGCTTCTCAC'GATC  
GACGCCCTCGACCACGACCTCTGCGCCCGCCAGTTTGAAGCAAAAATCCGCGGCGTCTGCGTG  
CTCGACCGCGTTCTGGCCGACCGCACGCTCGACTTCTGTCTGCTGATGTCTTCCATCTCCACC  
GTGCTCGGCGGCCTGGGCTATTTTCGGTTACGCCGCGGCCAACGCCTTCCTCGACGCCTTCGCC  
5 CAGGCGCGCAGCCCGCAGCCCGCTTTCCCTTGGCTTAGCGTGGCCTGGAGCGATTGGAAGTAC  
TGGACCGAGCGCAAGATGGACAACGAGGTCCGGCGCGTCATCGACAGCCTCTCGATGGAACCC  
GCCGAGGGCTTGAAGCCGTCACCCGCGTCTTGGCTTGGGGCAAGGCGCCCCACATCGCCAAC  
TCGCCCCGTGACCTCGGTCCGCCCGGGGATCAATGGGTCAAATGGCCAGCCTGAAATCGGCG  
CACTCCAGCGAGCCCGAGCCGGCTAGGCATGGACGTCCGGCGCTCTCCAGCGAATGGGTCCGCG  
10 CCGCGCAACGTGGTTCGAAGAGAAGCTGGTCGCCATTTTCGAGCAGGTGTTCCGGCACTGCGGCA  
CTGGGCATCGAGGACAAC'TCTTTGAGTTGCGCGGCGACTCGCTCAAGGCGGTTCATGACCGCG  
GCCCGTATTCAAAGGAGCTGAACGTGGAAGTGCCGCTGCCGACCTTCTTCCAGATGCCACG  
GTGCTGGCCTGGCCAGTTTCGTGACGAAGCCAAGCGCAGCGCCGGGAGACGATTCCGGCGC  
25 ACCGCGCCGCGCCACATTACCCGCTCTCGGCTGCCAGGGCCGCCATTACCTGCAC'TACCGC  
ATGGACCCGCGTTGTACCGCATAACAACGATCCCTTCGCCAACCTGATCGAGGGTCCGCTGGAC  
GTGGATCGCGTGGAGCGCATCCTGCACACCCTCATCTACGCCACGACTGCTTCCGCACCTCG  
30 TTCCACTTCCGCGAGGGCGAGCCGGTCCAGGTGATTACGATCGGGTGGACTTCAACCTGGCG  
CGGATTACCTGCGCGCCCGAGGATTTGCCCCGAACGGATGCGCGATTTCATCCGCTCCTTCGAT  
CTGGAGCGACCGCCCGCCATGCGCGCCGGCCTCTTCGTACGGGGCCCGAGCGCCACGTGCTG  
20 CTAATCGATTTTACCACATTATCACCGATGGCGTGTGTTTCGAGAACTTCGTGCGCGAGTTTC  
35 GCGGCGCTCTACCGCGCGAGATCCTGCCCGAGCTGGAACTCGAGTACAAGGATTTCCGCGGTG  
TGGCAGCATGAGAACCGGGGCCCGCCGCAACAGCGACCAGGCCCGCTACTGGACCGAGCAG  
TTGGCCAATGCGCCCGGGCCGATCGAGCTAACCACCGATTTCCCCCGTCCAGTCGACGCGAGC  
40 TTCCGCGGCGACCGCGTGCAGGACCGTGTGCTTGCAGGAGCTCGTTGCTCGACTCAAAGAGCAC  
25 GCGGCGCGCCTCGGCATCACCTCTATAGCCTGCTGCTGGGCGGATTCTCGTTATTGCAGCAC  
AAGCTCTCCGACTCGCACGACATCGTCATCGGTTCCGCCGTCCGCGGCCGACCCGGAGCGAA  
CTCCAGGATCTGCTGGGCGCGTTTCGTCAACACCCTGCCGATGCGCCACCGCATCGACCCGACC  
45 CATACCGCACGGGTCTTCTTGGAGCAGGTCCACCAGACAACCTTGGCGGCCCTCAGTACCAG  
GAGCACCTTTTTCGAGAAATGGTGGCGACGCTCGGTTTCGCCCGCGATCCGGCTCGCAACCCG  
30 ATCTTCGACACGATGTTCTTGTGTCAGAACATGCCCATGGGTGCAACCACCATTTCCCGGTCTG

5

10

CGGCTCTCGCCTCACGACACTTTTCACCGCAAGGCATTGTGCGACCTGATGCTACAGGCGACC  
GAGTATGACTGCCACCTGGAGCTGGTGCTCGAGTTCGCCACCGACCTGTTCCGGCTGGAACCC  
GCGCAAGTCTTGCTCGACCGCTACCGCCAAGTCTTGAGTGGCTGTTGGCGTACCCCCATGAA  
TCGATAGACGATTTGACGCTCGCCGGCCACTTTCGCGAAGTCGAAGTGACGATGTGGACGAG  
5 GGCGACTTTGATTTCTCAGATTTCAACCCCGCAACGTGAGAAACCTATGGCGCGCC

15

(2) peptide sequence

Seq ID No 58 (>pEPOcos6\_ORF9.pep)

20

10

25

15

30

35

40

25

45

30

50

55

MKLNVVANRLFDPESPERTEPAKSLLLAVTKVLPQEVNVRTRAI SVDLDRSFDAAAPAWAAS  
LLVECGAPVEETVVTYHGAARWLRRFDRVAVNGLGPFHPDQAPALLRERGVIYITGGLGGVAG  
QLARYLARACRARLVLTARRPLPERDQWDRESAVLSWDDKTRQRIELVRELERLGAEVLVVA  
DVADEAAMAQAIEASLARFDALDGLIHGAGIVRVASGRTPIGSMTRAMCEEQLRPKMLGLD  
DRLLRDRRLDFRIAISSSLAPILGGLGHVAYAAANLYMDAFATRAAAGNAPWIALNLAEWEYEG  
PATYDERVGRSLKQLELTNEEGIRVFQTVLALAARGPLQIIISTGDLQARLDKWIHIKSLHR  
RPGPVQLSRRTAAPQGGFGSERAFAAFADAWCDFGVEEVDPNKNFFDLGASSLDFIHLVS  
RFSKAIEQHVPLEALLEHSTLHDLAAHLAGDANTDASDEAIRQRLQGA KSGDIAIIGMAGRF  
PLAPDLDTYWRNLVGGIDAVSFFSAEELRAAGVTAAEIHHTNYVPAGRCADQDLDFDAFF  
TASDAELMDPQNRVLHEVVWHALEDACDFDNGDHQVGLFAGASPNLWWQFVASFSEAAKTQ  
MFTTTLLNDKDSIATQISYKLGLKGPVATLFTGCSTSLVAVDAACRSIWSGQSDMAVAGAVSL  
20 TLPDKAGYIYEKGMFLSADGHCFRANATGMVFGDGAGAVLKPLDAALRDGDPHIAV  
ATNNDGDRKAGYTSVSAQCQAEVIRSAQILADVAPESISYVEAHGTGTLGDSIEIKALKQAF  
ASDKNGFCGIGSVKTNLGHLMAGMAGLIKTVLAMKHRQLPPSLHCDEVNPDLELERSFFYI  
NTRLRDWVAPGGPLRAGVSSFGIGGTNAHVILEEPPTRESGTRMRHWKLLMLSAASEALDRQ  
ADNLADYLERHPEAHLSDVAYSLQTGRVRLAWRRTVLCEYREDAVTSRERQAKRVQTSRV  
40 DHKDVVFMFPQGAQYLNMGRLDYVMEPVFREVMDCFELLAPLWSEHPRQILYPEGGVSTLL  
HRTDYTPQIVFCFEYALAHLLLSWGLKPAATIGYSFGEYVSACLAGVFSLEDAIRLVTERGRL  
MAALPAGAMLSVPVPECELLRLLDGFHAQSAHLALAVDNGASCIVAGEQAAISAFESMLRKK  
RLLTMRVAVSHAHSQVMTGATDALRSILRKIPLSAPTIPFISCVTGTWITAQQATDREYWN  
HMCCTVRFAAGLTELQONREAVFLEVGPGRDLTLLAHRILADSAAVFELVKAPDGGDDDGFL  
50 LDRLAKLWRLGISIDWAGFYADERRRXLSPGYPFERRRFWIEGNPLEIAAGRPNVQGPLVKA

5

10

15

20

25

15

30

20

35

SDIGAWFYVPQWRRSVLAEPGTTAAGAAVTAEQARVVTEL RAGCASAGLGSGACGLNGGAPSE  
RPKESVAPAGSTSAAAQGTGADCTPTGEPAAVPKDGAEPRPTWLI FADAGGLAESFAKRVQAR  
CEKLYLVASGSRFERLAETRFRLDPGAKSDHRLLFKALDEADILPTHLLDFRSLDCGGPDADP  
MDQAGFFGLLHLVQAMAEAGYSHPIRLLI VSCGVYDVTGAEPLQPARATMIGPALCIPQQYPH  
LETSVDLGVVHADELHAAHQDLSLLAECLSAERQLALRGRHRWLLDYEPVRLPPLDPGRL  
PWRQRGVYLITGGLGGIGRILA EHLARTTSARLVLIGRETLPDRDDWD AWLNRQPVDATHER  
LLHKIRAIRDLEALGAEVLVLAADVANEAMREAYDRAESHFGTIHGVIHGAGLMDAQSFSLI  
DALDHDLCARQFEAKIRGVCVLDRLADRTLDFCLLMSSISTVLGGLGYFGYAAANAFLDafa  
QARSRDAAFPWLSVAWSWDWKYWTERKMDNEVGAVIDSLSEPAEGFEAVTRVLAWGKAPHIAN  
SPGDLGRRRDQWVKLASLSAHSSEPEPARHGRPALSSSEWVAPRNVVEEKLVAIFEQVFGTAA  
LGIEDNFFELRGDSLKAVMTAARIQKELNVEVPLPTFFQMPTVAGLAQFVTQAKRSGRETIRR  
TAPRPHYPLSAAQGRHYLHYRMDPRACTAYNDPFANLIEGPLDVDRVERILHTLILRHDCFRTS  
FHFREGEPVQVIHDRVDFNLARITCAPEDLPERMRDFIRSFDLERPPAMRAGLFVTGPERHVL  
LIDFHIIITDGVSFENFVGFEAALYRGEILPELELEYKDFAVWQHENRGRRANSQARYWTEQ  
LANAPGPIELTTDFPRPSRRSFRGDRVRTVLDAELVARLKEHAARLGITLYSLLGGFSLLOH  
KLSDSHDIVIGSPVAGRTRSELQDLLGAFVNTLPMRHRIDPHTHTARVFLEQVHQTTLAALSQY  
EHPFDEMVA TLGFAADPARNP I FDTMFLQNAMGATTI PGLRLSPHDTFHRKALCDLMLQAT  
EYDCHLELVLEFATDLFRLETAQVLLDRYRQVLEWLLAYPHESIDDLTLAGHFREVEVTMSDE  
GDFDFSDFEPRNVRLWRA

pEP0cos6\_ORF10 sequences:

(1) nucleotide sequence

Seq ID No 59 (>pEP0cos6\_ORF10.seq)

ATGGCGCGCCTGAGCCGCACAGATCTCCA ACTCGCCATTCACCAGCGCACCGTGGAGCGCGAA  
TATTGGCGCGCTCTGTTCGAGCGCCATCCGCAACGGTCCAGTTTGCCGGGGGTGCTCACC GCC  
CCGATCGGCGACGAGTCGACCCGCGAGACCTTGTCATTTCGTCCTCGACGAAGATCCCCTTCGG  
CTGAGTAATCGTTCCCGCAACGCCTGCTCACGGTGTGGCGGCTGGCCTCGCGGCTTTCTC  
CACCGCTGCGACGGCGCTGAGCGCTTCA CCCTGGGGTTGGCCCTACCGCGCCAAGCCGATGAC  
CATCACCCGATCCTCAACAGCTTGATCGCGCTGGGGGTGCGGCTCGACTCGAGTACGACCTTC

50

55

5

10

15

20

25

30

35

40

45

50

55

CGCGATCTGCTCTATGCGCTTCGATCCGAATACCACGAGGCGATGCGCCACGCCAACTTTCCG  
CTGGCGACCTGGTGGCGCGGCCTACCCGGCGGAACGGCGCCGTTTCGACGTGCCCCCTCAGCCTG  
GACCCCTTCACAGACGGCGATTTCGCTGGAAGACCACGCGATCGGCGCGTTGTTCCGGTTTCGA  
TTGGAGGGTGAGCGCCTCACCTGCCGATTGCGATTTCGACCCTGCGCGCTATGACCGTCCC GCG  
5 ATCGAAAACCTCGCCGATCGTTTCGCCCCGCTTCCTCACGCGCCTGTGCCGGGACGCCTCCACC  
15 GTCATCCAGGCGCTGGACCTTTCGCTGCCAAGCGATGAATCGGTGTGGCGCGTCACTGAAGGC  
GTGCGGCGCGGCTATTTCGAAGACCTGACGCTAGACCGCGCGTTCCGCCGCCAGGCCGCGCAA  
ACGCCCCGATCAGCCGGCGATCACGTTGAACGGGGACGTCCAGAGCTACGCCGAGGTTCGACCGC  
CGCAGCGACGCGCTGGCCCCGCCACCTCCGTCGCCACGGCGTCGGTCCGGAAACGATTGTGGCC  
20 GTCAACCCCCGGCGCGGCCTAATCAGCTGACGGCCCTGCTCGCGGTCCATAAGGCCGGCGGC  
GCCTACCTGCCGATCGATGCCGAGGAGCCGGCTGCCCGCCAGCAATTCAAGGTGCGCGACAGC  
GGGCGCGGTTGGCACTGGAGCCGTGCGCGGACCAGGCGCTGACCGTCACCGACCTGCCGCGG  
CTCTTCTTGACGATGCCTCGCTCTTCGCTGACGGCGGGCTCGATGTGCCGCGCGGCGCGAC  
25 TCGCTCAATCCGGCCTATGTGATGTACAGTCCGGCTCGACCGACAGCCCAAGGGTGTGGTG  
15 GTTCCCCACCGCGCGTGGTCAATCGTTTGAATTGGGGGCGAGTCCCGTTTCCCGCTGGACGAA  
CGCGACCGAATCCTCCAAAAGACGCCGTGCTGTTTCGACGTGTCCGTTCTACGAGCTGTTCTGG  
30 GGCGCATGGAGCGGGGCCACCTGGACATCCTCGAGCCCGCGCCGAGCGCGACCCCGACGCA  
GTGGCCAGGGCCCTGGCCGAGCGCGCCATTACCGTATGCCATTTCTGTCCTTCGATGCTGCTC  
GTCTACTTGGAAGTCATGCGGCGGCACCATGCGCCGCCCGTGGCCGACCGGCTCCGTTACGTC  
20 TTCGTGAGTGGCGAGGCCCTCGAACCAGGACACCTCGCCGGGCTCCAGCAGATTGGTCGGCGC  
35 CTCGGCCGCACGATTCCCCTCGTTAATCTGTATGGACCAACCGAGGCCTCGATCGAAGTCTCC  
TGCTTCGCCTGTCCCGCCGACCATGTGCCGCGCCGGATCCCCATCGGGCAGCCGATCGACAAC  
GTCGCACTGCACGTTCTCGACCGGCGCGGCCGTGCCAGCCGCCCTATCTTCTTGCGGAGCTG  
40 TTCTTGCCCGGCGACTGCCTGGCGCGCGGTACCTCAACCGTCCCGACCTGACCGCGCTCCAC  
25 TTCGTGCCCAATCCCTTCGGCAACGGCGAGCGCATGTACCACAGCGCGGACTTGGCGCTCGTG  
CGCGCGACGGCCAAGTGGCGTTTCTCGGCCGCCGTGACCACCAATCAAAATCCGTGGTCAA  
CGGGTCGAAC TGCGCGAAATCGAGAGTCATTTGCGCGGGCTCGAAGGCATCGCCGCCCGGTC  
45 GTCCAGGCCGAGTCGCAGCACCATGAAACCCTGCTGCACGCCTACGTCGTCACCAACGACGCG  
GGCCTCAATGCGGCCCGGCTGCGCGCCGCCCTCGCTCAACATCTGCCCGAGTACATGATTCCC  
30 CAGCGCTTCTCGCGGCTGGCCGAGTTGCCGCTGCTGGCGGCAGGCAAGATCGACCGCGCCGCC

5

125

10

CTCGCGCAACGTGCAACGCCGCTCGCCAGCGGCGGCCCCCTTCGTGGAACCCAGCGGGCCCCACC  
CAGCAGCGTATCGCAGAACTGTGGCGCCAGGTCTTAGCGGTCGCCGAAGTCGGCGCCGAGGAT  
CCCTTCTTCAGCATCGGCGGCAACTCGCTCAATGTGCTCAAGCTCAGCGCCGCGCTGAGCGAC  
GCCTTCGCGCGTGACATTCCTCATGCCGGCCCTGTTCCAATACGACACCATCGCCGCCAGGCC  
5 TCCTGGCTCGACGGGCAGGTTGACGAACGGGCCAATCCGCCGCGCTCGACCGGCAGGCCGCC  
15 GAGGCGGCGCTGACCCTTCAAGAGACCGTGGCCATTTTGGAGGGATTGATGACGAACCA

15

(2) peptide sequence

Seq ID No 60 (>pEPOcos6\_ORF10.pep)

20

10 MARLSRTDLQLAIHQRTVEREYWRALFERHPQRSSLPGVLTAPIGDESTRETLSFVLDEDPLR  
LSNRSPQRLLTVLAAGLAAFLHRCDGAERFTLGLALPRQADDHHPILNSLIALGVAVDSSTTF  
RDLLYALRSEYHEAMRHANFPLATWWRGLPGGTAPFDVALSLDPFTDGDSDLHAIGALFRFA  
25 LEGERLTCRLRFDPAFYDRPAIENLADRFAFLTRLCDASTVIQALDLSLPSDESVMRVTEG  
VRRGYSQDLTLDRAFRRQAAQTPDQPAITLNGDVQSYAEVDRRSDALARHLRRHGVGPETIVA  
15 VNARRGPNQLTALLAVHKAGGAYLPIDAEPAARQQFKVRDSGARLALEPSPDQALTVDLPR  
LFLDDASLFADGGLDVPRGADSLNPAYVMYTSGSTGQPKGVVPHRGVNVNRLNWQSRFPLDE  
30 RDRI LQKTPLLFDVSVYELFWGAWSGATLDILEPGAERDPDAVARALAERAITVCHFVPSMLL  
VYLEVMRRHHAPPVPDRLRYVFSGEALEPDHLAQLQIGRRLGRTIPLVNLGPTASIEVS  
CFACPADHVPRRIPIGQPIDNVALHVLDRRGRROPYPYLPGEFLAGDCLARGYLNRPDLTALH  
20 FVPNPFNGNRMYSGLDALVRGDGQVAFGLRRDHQIKIRGQRVELGEIESHLRGLEGIAAAV  
35 VQAESQHHETLLHAYVVTNDAGLNAAARLRAALAQLPEYMIQRFSLAELPLLAAGKIDRAA  
LAQRATPLASGAPFVEPSGPTQQRIAELWRQVLAVAEVGAEDPFFSIGGNSLNVKLKLSAALSD  
AFARDIPMPALFQYDTIAAQASWLDGQVDERAQSAALDRQAEEAALTQETVAIFEGFDDEP

40

25

45

50

55

5

126

## pEPOcos6\_ORF11 sequences:

10

## (1) nucleotide sequence

Seq ID No 61 (&gt;pEPOcos6\_ORF11.seq)

15

20

25

30

35

40

45

50

55

5 ATGACGAACCATGACCATCACGAGGAGAGCAGCGGCCTGGAGATCGCCGTCATCAGCATGGCC  
TGCCGATTCCCGGGTGCTATGGCCTGCCGATTCCCGGGTGCTGCCGATTGCGACGCATTCTGG  
GAAAACCTGATCAACGGGACCTCCTCGATCACCCATTTCAGCCACGACGAGCTGATCGCGGCC  
GGCGTTGACGCGCGCGACCTGACGCGCAGTACGTGCGCGCGGCCGCCAGATCGATGACGCC  
GAACGGTTCGACGCGGCCCTTCTTTGGGTACTCCAGCGTGAGGCCGAGCTGATGGACCCCCAG  
10 TTCCGCCTGCTCCATGAATGCGCCTGGTCCTGTCTGGAACAGGCCGGCATCGATCCGCGCGTC  
GAAGCCGCGCCGATCGGGCTGTATGCCGGCGCAGCCGACAACACCTACTGGAACGCGCTCTCG  
TCGCTCGACCGGGGCTCGGCCGAATCGGAGCAATTGCGCGCGGAACAACCTTGCAACCGCGAT  
TTTCTGTGACGCTGGTCGCGCGCGCTCAACCTGAAAGCCCCGCGGTGGTGGTTCAAAGC  
GCCTGTTGACCTCGCTGTGGCGGTCCACTCGGCCTGTCTGCGCTCCTGACCGGCGAATGC  
15 CGAGTGGCCTTGCGCGGTGGGGTGGCGCTGCGCTTCCCACGCCCCAGCGGTTATCGCTACGAA  
CCTGGCATGATCTTCTCGCCCGACGGGGTGTGCCGCGCGTTCGACGCGGGCGCTAACGGGACG  
GTGCCCCGCCAAGGCGCGGGGCTGCTAGCGTTGAAGACGCTGAAACGTGCCCTCCAGGACGGC  
GACACGATCCACGCCGTGATTGCGCGGACCGCGGCAACAACGATGGTGCCCGCAAGACCGGG  
TTCACGCGCCCCAGCGCCACGGCCAAGCCGAAGTCATTGCGACGGCGCTGCGCCTGGCCCCG  
20 GTGCCGCGCGAATCGATCGACTACGTGAGGCCCCACGGAACCGGCACGCCGCTAGGCGACCCG  
ATCGAGGTAGCCGGCTTGGTGGAGGCCCTTCGCCAGCGAGAAGCGCGGCTATTGCCGGCTGGGC  
TCGGTCAAATCCAACCTTGGTCATCTGGACACTGCTGCCGGCATCGCCGGCCTGATCAAGACC  
GTGCTGGCGCTCGAGCACGCGCACATCCCCAAGTCCTGCCACGTGCCACGCCCAACCCCGCG  
40 GCGCGCTACACAAGACGCCTTTCCGCATTGCCGCGACGGGATGGCCTGGCCGCGGCGTATG  
25 GCGACGCCGCGGGCGGGCGGGTGAAGTTCGTTCCGGCATCGGCGGCACCAACGTCCACGCGATT  
TTGGAGGAGCGCGCCCGCGCGCGCGGAGCTGGCGGACGGGCGCAGTCAGGTGTTCTCTTC  
TCCGCCAAGGACGAGGCGGCGCTGGACCGTGCCCTTGCCAACTATGGTGCGGCCTTGGAGAAG  
CGCGGCGACCTCGCGGCGGGCGCGGTGGCTGGACGCTCCAAAACGGCCGGGCGCATTCGAA  
TGCGGAGCCAGCGCGGTGGCATCCGACCTCGACGAATTGGCGGGCGCATTCGCGGGCGAGCGG  
30 CCCGGCGCGTCAAGAAAAACCGAATGGCGCGGAGGATAAGCCGGTGGCGTTCTTATGTTG

5

127

10

5

15

20

10

25

15

30

35

40

25

45

30

50

55

GGGCAGGGGAGCCAGTACCGTGGCATGGGCCACGACCTGTACCGCGAAGAGCCGCGTTTCCGG  
CACCACCTCGACGCCTGCCTCGCCATCCTCGCCGAACACAAGCCCCGAGATCGACTGGCTGGCG  
TTGCTGGGCTACCGCGACGAGGACGAGCCAACCGACCAGATCGGGACGTCCCTCGCAGGGCCCG  
AGCCGGTCAGCCGCATCGAACCCAGCGGAGCTCCTCGACAGCACCGAATTGCCCCAACCTTTG  
CTTTTCTCCATGTCCTACGCGCTCGGTGCGCTGTGGCTCGACTGGGGCGTGCGACCCACGGCG  
ATGATCGGGCACAGCCTGGGCGAGTACAGTGTGTCATGTATTGCAGATTCTATGCACTCGAT  
CAGGTGCTGCCCTTCATTCTGACCCGCGGTGAGTCATGGCGCAATTGCGGCGCGGCTCGATG  
TTGGCCGTGACGGTGACAGCGTTCTGATGCGCGAGCTGATCGCCGATGCGCTCGATTGCGG  
GCGATCAACGGCGCTGACCAATTTGTCTGGAGCGGGCCGAGCGAGGCTGTCCAAGCCGCGGGG  
GTCCGACTGCGCGGCGCCGCGCTGCGTGCCACCGAGCTGAACACCTCACACGCGTTCCATTCA  
GCCATGATGGATCCCATTTCTGGAGSAGCTAACGGTTGCCGGTTCGCGACTTCAGGTGCGTGT  
GGGACGATTCCGGTCGTTTCATGCGTTACCGGAACCTGGTTGACGGCGAAGCAGCTGGCCGAT  
CCGCGCTACCACGCGCTCACGCGCGGAACCGGTGCGGTTGCGGCGGGCCTAGCGACGCTG  
ACAGGGGAGGAGCCCGCTGATGCTCGAAGTGGGGCCGGGCTCGACCTGGCGGCTTTGGCC  
CGCGAGCATTGCAATGCCCCGCTCCCGGTGCTCACCAGCCTGCGCCACGCTCGCCAGGCGACG  
CCCGATCGCCAATACCTGCTCGAAACGCTCGGCTGCCTTTGGCGACACGGGTTTCCGTCGAT  
TGGGGGGCCCATGCCGACGTTCCGCGACGCTTGTTTCGCTGCCCGGCTATCCCTTTTCCGGC  
GCGGTGCGCCGCTTAGCCGGCGACCCCTCCGCTGCTGGCCGAGCCCGCGCCGTCGCCGCC  
CCGTGCGGAACGCGCCAATCAGCGCCGACGCGCGGACCTCCCGAACACTCCGAGCCGACA  
TCCGGCGCCGTGTCGGCGATCAAAGCGCCAATCGCCGCCGCGATCCCGGCCTCTATCGCCTC  
TCCTGGCGCCAGGCCGGAACGGCGCCGCTCGGTCCGCCGATCTCGGTCCGCCCCGCGACTGG  
ATCGTCTTCGCTCTGATTCTCACCTGCTCCAGGCGCTCAGGGCCAATCTCGGGACGCGCGCT  
CAGCGGGTGACGCTGGTGACGCCGGGCCAGGAGTACGCAGCCGAGCCGTCCGGGTTTCGGCTG  
CGGCCGGACCAGATCGACGATTACCGCGCCCTGTGGGCGGACTTGGCGCAAACCGGTATTGTG  
CCACGATACATCGCGTTCCTCGCCCCGTTTCATGTACCGGGCGCGCATGGCGGGCGATGCCTCG  
ACCCTGGACGAAGTGCGCGAGGGCGGCTTCCTGCCCCTGACCCGCTTGATCCAGACTCGCCCCG  
CCAGGGCGACCGAGCGGACTTCTAAGCCTCACGATCGTCACCCCGCGCCCTGGCGCTGGGC  
GACGAAGCGACGCGCCCGGAATGGGCAATCCTGCACGGGATGGTCGCCGGCTTAAGCCGCGAT  
TATCCCGAATGGCGCTTCGTCTCGATCGACGGCGCGACCCATCCCCGCATCGGTGCGAAGGT  
CTGCCCCGCTTGATCGCGCTTCATGCGGTGACGAGGCTGGCCCCGACCCGCTTGGCGCTGCGC



5

10

15

20

25

30

GGCCTTCACGCTTGGGTTCACAGTGCGAGCACGTTACGCCGGCCACCATCCCTGGGGCGGGT  
ATGTGGCGCGAGGGTGGTGTGTACATGATAACGGGCGGATTCTGGCGGGATCGGTCTGGCGCTG  
GCCCCGCGCCCTGGCTCGAGAAGCTCGCGCCAAGCTGATCCTGGTCGGCCGAAACCTGCCCACC  
CGCCCGATCGATCTCGAGGCTTGGGACGCGCGCCGCTTGATTCTCACCGCCGACGTGCGCGAC  
5 GAAGAGGCCATGCGCCGCGTCTTCGATGCCGCGCACGCCCGGTTCTGGCGCCATCGACGGCATT  
CTTCACGCGGCGGCTCTCCCGGTGGCAGCCTGTTCCGCCAACCAATCGGACGCGGCCTTCGAA  
15 GACGTGCTGCACGCCAAGGTTTCGCGGTACCCTCGTGCTGCAAGGCCTGAGGGCAATCGATGCG  
CCGTGTTGCTGATGTCCTCGCTGGACGCGCTGGCTTCCCGGTCCCGGTGACACCGCTATGCC  
GCCGCCAACGCCTTCTCGACGCCTTCGCCAGTCTGCGCCGGCGAGAGGGAGAGCCGGTGATC  
10 AGCGTTGGCTGGGACAGTTGGTGCAGGTGGGCATGGCTGCTCGGGTCGCTGCCCGATCGGCC  
GACGAACGCGGCCGCTTGGCGCGCGAGGGGATCAGCCCTCGCCAGGCTTGGCAGGCTTTGAGC  
CGGGCGCTCGCCCTCGACCCCCCCCCACCTGATGATCTCGCGCACCGACCTGACCTCGCGCTGG  
25 CACAGTCGATCCAGCCCTACGCCGCTCGCCTCGAGCGAACCCGAGGTGGCGCTGCCGCGCTGG  
ACCGCATCCGCCTGCCAAGCCGTATCGAGCGTGTGTTGGTGCAGCACTTCGCCACCGCCGCC  
15 GTGCCTCCCGATGGCAACTTTTTTCGAGCTCGGCGCCAGTTTCTTCGACATCGTCCAGCTCAGC  
GCTCGACTTCAACAACAGTTTCGGCCGAGATGTCAGCCACACCGTGCTCTACAGTCATCCACC  
30 GTCGCCTTGTGGCCGGCTACTTCGCCAATGACCCGACGCCGTCCGGTGCTGCTGCCGACGAA  
CGCGACGAAGCGGTGCGTTCGCGGCCGCGACCTCTTGAAGAGCCCGCGGCGAGGAGTA

35

40

45

20 (2) peptide sequence  
Seq ID No 62 (>pEPOcos6\_ORF11.pep)  
MTNHDHHEESSGLEIAVISMACRFPGAADCDAFWENLINTSSITHFSDDDELIAAGVDARDLT  
PQYVRAAGQIDDAERFDAAFFGYSQREAEELMDPQFRLLHECAWSCLEQAGIDPRVEAAPIGLY  
40 AGAADNTYWNALSSSLDRGSAESEQFAAEQLCNRDFLCTLVAAALNLKGPVVVQSACSTSLLA  
25 VHSACRALLTGECRVALAGGVALRFPRPSGYRYEFGMIFSPDGVCRPFDAGANGTVPGEAGL  
VALKTLKRALQDGDTHAVIRATAANNDGARKTGFTAPSAHQAEVIRTALRLARVPAESIDY  
VEAHGTGTPLGDPTEVAGLVEAFASEKRGYCRLGSVKSNLGHLDTAAGIAGLIKTVLALAH  
45 IPKSCHVATPNPAARLHKTPFRIAADGMAWPRRMAPRRAVSSFGIGGTNVHAILLEEAPRA  
PELADGRSQVFVFSKDEAALDRALANYGAALKRGDLAAGAVAWTLQNGRAAFEWRAVAS  
30 DLDELALALRGERPGAVKKNRMAREDKPVAFLCSGQGSQYRGMGHDLYREEPRFRHHLDACLA

50

55

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55

ILAEHKPEIDWLALLGYRDEDEPTDQIGTSSQGPPSRSAASNPAELLDSTFAQPLLFSMSYAL  
GRLWLDWGVPTAMIGHSLGEYSAACIADFYALDQVLPFILTRGRVMAQLRRGSM LAVSGDSV  
LMRELIADALDLAAINGADQFVWSPSEAVQAAGVRLRGAGLRATELNTSHAFHSAMMDPILE  
ELTVAGSRLQVGVGTIPVVSCVTGTWLTAKQLADPRYHARHAREPVRFAAGLATLTGEEPFLM  
5 LEVGPGSTLAALAREHSNARLPVVTSLRHARQATPDQYLLLETGLCLRHWGVSDWGAHAGRS  
10 RRLVSLPGYPFSGAVRRLAGDPLRLLAGARAVAAPSGTRQLSADARDLPNTPEPTSGAVSAIK  
15 APIAAADPGLYRLSWRQAGTAPLGPDLGPPRDWIVFASDSHLLQALRANLGTRAQRVTLVTP  
GQEYAAEPSGFRLRPDQIDDYRALWADLAQTGIVPRYIAFLAPFMYRARMAGDASTLDEVREG  
20 GFLPLTRLIQTRPPGGPSGLSLTIVTPAALALGDEATRPEWAILHGMVAGLSRDYPEWRVVS  
10 IDGGDPSPHRCEGLARLIALHAVDEAGPTRLALRGLHAWVPQCEHVQPATIPGAGMWREGGVY  
MITGGFGGIGLALARALAREARAKLILVGRNLPTAPIDLEAWDAPPLILTADVADDEAMRRVF  
25 DAAHARFGAIDGILHAAGVPGGSLFANQSDAAFEDVLHAKVRGTLVLQGLRAIDAPLLLMSL  
DAWLPGPGQTAYAAANAFDLAFASLRRREGEVPVYSGWDSWCEVGMAARVAARSADERGRLAR  
EGISPRQGWQALSRLALDPPHLMISRTDLTSRWHSRSSPTPVASSEPEVALPRWTASACQAV  
15 IERVWCEHFATAAVPPDGNFFELGASSFDIVQLSARLQQQFGRDVSHTVLYSHPTVALLAGYF  
ANDPTPSGAAADERDEAVRRGRDLLKSRRRGV

pEPOcos6\_ORF12 sequences:

20 (1) nucleotide sequence

35 Seq ID No 63 (>pEPOcos6\_ORF12.seq)  
ATGACCGTGGAGCACGAAACCGGATTCGAAATCGCCGTCATCGGGCTGGCTTGCCGCGTTCCC  
GGCGCTGCCGACGTGGCCGCCTTCTGGCGCAACCTGGTCGAGGCCAAGGAGAGCGTGCGCTTC  
40 TTCGAGGACCACGAGCTGCGGGCCCGCGCGTCCCGAGGAGATCTTGCGCCTGCCCAACTAC  
25 GTGAAGGCCAAGCCACTGCTCGCTGATGGCGAAGCTTTCGACGCGGACTTCTTCGGGTTCCAT  
CCGCGCGAGGCCGCTACCTGGACCCGCAAGTTCGGCTCCTGCACGAATGTTGTTGGACCGCG  
CTGGAGGATGCCGGCTACGATCCCGCGCAGTACGCCTACCCGATCGGGTTGTTCCGGGGCGTC  
45 TCCAGCAATCTCTCGTTCCTGTTTCGACCGCATCGATCCGCGCGACTCCCCCTGCAGAAGCGC  
TATGTGGCCGAGCTGAACGCGGCCTCCTTCGCCACCCAGATCGCCTACCGGCTCGATCTGAAG  
30 GGGCCGGCCATTTCGATTCAAACCGCCTGTTTCGACGTCACTGGTGGCGATTACCTGGCGGCG

5

130

10

CAAAGCCTGATCGGCGGCGAGTGCCACATGGCCTTGGCCGGCGGAGCGACCTTGGAGGTCCCC  
AAAAAGCCCCGGCTATCTCTACCGCGAAGGCTACATCAACTCGCCGGACGGCCACTGCCGGGCC  
TTCGACGCCGACGCGGCCGGCACCATCTTCGGCGACGGCGTCGGCATCGTCCTGCTCAAACGC  
TACCGCGACGCCCTACGCGACGGCGATCACGTGTACGCAGTGATCAAAGGCTCGGCCATCAAC  
5 AGTGACGGCCATCGCAAGGTGTCTTACACGGCGCCGGGCAAGAGCGGTCAAGTGGCGGTGATC  
15 CGCGCTGCCTGGCGGCGGCCAGGTAGAGCCGCAAACCATTCGCTTCGTCGAGGCCCCACGGG  
ACCGGCACACTCGCCGGCGATCCGATCGAGGTAGAGGCGTTGACGGAGGTCTTTGCCGAAGCG  
GGTCGCGGTACCTGCGCCCTGGGTTCGGTGAAGACCAACATCGGCCACTTGGATGTGGCGGCG  
GGCGTGGCCGGTTTCATCAAGGCGGTCTTGGCGCTCGAGCGGCGCTCCTCCCGCCAGCCTT  
20 CACTTCGTCCGGCCCAACCCGGCCATCGATTTCACGGGCCCTTCTACGTTTGTTCGCCAAATC  
10 GAGCGGTTGACGGAGAACGGGCGGTTCGGGCGCGGGGTGAGTTCCTTTGGCATTGGCGGCACC  
AATGCCACCGTGATTCTGGAGGAAGCGCCGGCGCGGAGGCGAGACTGCCGGCCGGGAGCCCCG  
CCAGGCGCGAGTCCGTTCTGTTCCTGCTATCGGCCAAGACGCCGATGCGCTGGCAGGCCGT  
25 TGCCACGACCTTGCCGACCACCTGCGGGCGCACCCCGAGCTCCTCCTGGCCGATGTGGCCCTC  
15 ACTCTGCAGATGGGGCGGGCGTCTTCGCCTACCGCCATGTGGTCCAGGCTGCGACGGCGGAG  
GAGCTGATTTCGCGGTCTGGGAGCGTTCCGACAGGAGTCCATCCGCAAGAGGCGGAATCGAGTA  
30 CAATGGGTGTTGGCAGGCGAGGCGATGTGCTTGACGCCGGTTTGGCGCTGTACGCCGATTGG  
CCGGTCTATCGGGAGCGGGTCGACGTCTGTCTGGCGATCGTCGCCAAGCTGCGCCAAATCGAC  
GGCCGGTCATTCTACATGAGTGAATCGAGCGACCGCGCGAGGTTCTGCCGAATGGTCGACG  
20 GCGCTGGCGTTTCATGTTCCACTGCGCGCTGGCGCAAGCCCTGAGCCAGGCCGGCCTGCACCCG  
35 CAGCGCATGTGGAGCCGTGGGCTGGGCGGACAGGTGCGCGTGGTTTTGGCCGAATCCCTGTCTG  
TTGGAACAAGCGCTGGCGCTGGTGTTGTGCCAGACACCGGTTCCCGGCGATGCCACACCTCAG  
CGCGAACGCTTGGTTCGGACACTGGAAGGCTGCCGGTTTCGTCCACCACGATTTTTGATTTCTG  
40 GCAGACAGCTCGGGTCGACCCCTGGACCTCGCCGAATTCGCTCATGTGATTTTTGGTGGCGGT  
25 GGCCAAAGCGCCTCGCCCAATGAGGCGGAGCTGCGCTCATGGAGCGACGCCGCGCCCGAGCTG  
GTGACCTTGGCGATCGGCCCATCTTTCTCGAGGCCGCTCCGGGACGGTGGGTCTGGCGATC  
GACCCCAAGCGACCGATGACCTGTGTTTCAGCGCACGGTGGCCGCGTTGTGGGAATGGGGATGT  
45 GACGTGCGCTGGGCTGCGTTACCTCGTCGACCGGGCGTGGGTTCCCTGCCTACCTATCCC  
TTCGTGCGGGTAATTCCCACGATCGGCGACCCCTTCGCGGAGCAGGCGCGGAGGATGACTTG  
30 ATTCGGGCGAGCGCTTCCGCGTCGSCCGGATCGCCGCCGAGCCGTCCGCAAACCTCGGCAGCG

50

55

5

131

10

15

20

25

30

35

40

45

50

55

GAACGCCCACGCGCCAGTCAAGCATCGCCTCGGCAACCACACCGGCTCCGTCTCATACGTGCG  
GCCAGCGTGGCCGTGGCCACCATTCTCGAAACCGTCCGTGCCTATTTGCGGTTCGCCGCCGTG  
CGTTCCACCGACGCCTTCTTCGAATTGGGCGCGTCTCGCTGGATTTGGTCAACCTGGGCCAG  
CTCCTTTCCGATCGTCTCGGCCGCGAGGTTCCGACCCTGCTCTCTACGACCACCCAACACCG  
5 GACCAGTTGGCGCTGGCCCTGACATCCGCGGCGCTCAGCGCAGAGGCGCCGCCCTTAAGGGGC  
GGTCATCGCGCATCGACTTCCGGCACAGCCGCGAGCTCGGCCGCTCCACCGCACCGACGTTT  
15 CCGGGGACGCTCACTCGAGCCAGCTTCGTTTCGCGAGCAGGACATCGCCATCATCGGGATG  
GCCTTCCGGGACCGGGCGCCGACGACCTGGACGCGTTCTGGAACAACCTGGTGAAGGGGTC  
GAGTCGATCACCTTCTTCAGCGAGGACGAGCTGCTGGCGGGCGGCGTCCCCCGGAACATCTG  
20 GCCTCGACGCGCTACGTGCGGGCCAAGGGGAACTGACTGGGATGATGGATTTGAACCGGAA  
TTTTTCGGTTATTTCGGCGCGGAGCGCGGTCATGGACCCGAGTTCGCGTGTTCACCGAA  
TGCTCTGGCACGCACTGGAGCACGGCGGTACGATCCGACCCGATGCGCGGCATCGATTGGC  
GTCTACCGCGCGTGACCAACCACCTGCCTTGGCTGATGCGAACTTTGCCGCACCTGACCGAG  
25 GAGGAGCAATTTCGGCGCGCTGCTCTCACCGACCGCGAGTTTTTCGCACCGCTGCTCTCCTAC  
AAGGTGGCCCTGCGCGGACCCGCTATTTTCGCTGCAAACCGCCTGTTGACGTCGTTGGTGGCG  
ATCGGCACGGCCTGTCGGAATTGCGCGCGGGTGCCTGTGAGATGGCCCTAGCGGGCGCGTG  
30 ACGGCCAGCATCGAGCGCTGCGGCTACTTCCACCAAGAAGGCTACATCCTCTCGCCTGACGGC  
CACACGCGCAGCTTCGACGCGGCGGCCCGGCACGGTCTTCGGCGACGGAGTCGGCATGGTG  
CTGTGAAGCCGCTGGCCCAAGCCTTGGCCGACGGCGACACGATCCACGGGTGATCAAGGGA  
20 ATCGGCATCAACAACGACGGCGCGCGCAAGGTTCGGCTTACCGCACCTAGCCGGGCCGGTCAG  
ACCGAGGCGATTTCGGCCCGCGCTGCGCGACGCGGGGTGGCGTCGAACCGCGTCAGCTACGTG  
GAGGCGCATGGAACCGCGACAGAAATGGGCGACCCGATCGAGGTCGAGGCCTTGACCCAAGCC  
TTTCGCGCCGAAGCCGACGGTCCGCTTCCGCCCCGCTCCTGCCTACTCGGCTCGGTGAAGTCC  
40 AACGTGGGCCACCTGAACGCCGCGGCGCGGCTGGTGGTCTGGTAAAAACCGTGTGGCGCTC  
25 CAACACCGCCGCTGCCGACCAGCCTGTTCTACAGTCGCCCAATCCACACATCGACTTTGCG  
GCGAGTCCGTTCCGCGTGAACGGCCAGACTTCGGATTGGGTGCGGCCAGAGGGGACGCGGTTG  
CTGGCGGGAGTGAGTTTCGTTTCGGTATCGGGGGAACCAACGCCACCTGATCGTCGAGGAGGCG  
45 CCGAAAGCGCTACCGACGACAGCGGCACCTCTGTGACGGAGCCGAATGACCTCGACGCGGGC  
GACGCCGACGGGCTAGTGCTGCCGATCTCGGCCCGCACGCCGACCGCCCTGGCGCACATCGCG  
30 ACCAACCTCGCCAATCACCTGGAACGACATCCGACCATCGCCCTGGCCGACGTCGCCCTGACC

5

132

10

15

20

25

30

35

40

45

50

55

CTTCAGCTGGGCGTCGCCAATGGCCCCATCGCCACAGCCTGATCTGCCGGAATCGAACGGAG  
GCGATCAAGCTGCTGCGCGCCGTCGTCCACTCCGCGGAGGTGCCGCCAGCTCAGGCGCCGGTC  
TCGGATGCGCCGCGCTGTGTTTTCTTTTCCCGGCCAGGGCGCCCAATACCCGAGCATGGCC  
CGCGACCTGGTTTCAAACCTGTCCCGACTTCGCCCTGCACCTGGACCCCTGCCTCGACCACTTG  
5 GCCGAACTGCTTCCCGAAGATCCGCGTTGCATCCTGTTTCGGCGATGGCCCCGCCGATCGGCTC  
15 GACCAGACGGCCTACACTCAGCCGCTGCTCTTCTCCGTGTCCTACGCCTTGGCGCGCTGTTG  
GGCGATTTCCGGCATTCGCCCCGATGCGATGATCGGCCACAGCCTGGGCGAATACGTGGCGGCC  
TGCTTGGCCGGGCTTTTCTCGCTGAGCGATGCCCTGCTGCTGGTGAGTGAACGCGGCCGCTG  
ATGGGCTCGGCCGCGCGGAGCGATGTGGCCGTCCCTTGGCCGAATGGGAACTGGAGGAA  
20 CGCCTGGAGCTTCTGGCCGACGACCGAATCAGCATCGCGCGGTCAACACCGCCGAGAGCTGC  
10 GTCATCGCGGGACCCAGCGAGGCGATCGAGCGCTGCGCCCAGCGCTGGGCCGCGCAAGCCCTG  
ACCTGTACGCCGCTGCGCACGTCCCACGCCCTTCCACTCCGCGATGATGGAGCCGATTGTGAA  
CCCTTCGSCCATGTCTTGGCACGGGTACCTTCGCGCGCGCCGCGCGCTGGATCTCGAAC  
25 CTCGACGGCAAGCCGATCGATTCCGCGCGGTGATGCAGCCGACTATTGGGTGCGCCACCTG  
15 CGCCAACCGGTCCGCTTTCACGAGGACTCAGTCACCTGTTGGCCGAGGACACCCATGCTTGG  
GTGCAAGTGGGTCCCGGCCGAACCTGTCTCTTCTCGTCGCGGCCACCCGGCTACCGTCAC  
30 CAGCCAATCGTCAACCCCATGCGCCATGCAGTCGAGTCGACGGGCGACGTGCGCCGGTGGCGC  
CAAGCGCTGGGCGAACTATGGCGGGCCGGCATGCCGGTCCCTGGGAGCGGCAGCGCGCGGC  
CGGCATGCCGGACGACGTGTGCCGCTGCCGGGCTACCCCTTCGAGCGCGGCCCTTCGCGGCC  
20 CGAAGACCGGTGGAGCTGGCGCAGCCCGCGCCCAAGGCGGAGCTGGTGA AAAAACCCGATCCC  
35 GCGCGGTGGCTGTACCGCCGCGTCTGGCGCCCTGCCAGGCTGCGGCCGGCGGACTGGCGGTG  
CAGGCGACCGTTCTGGTCTTCGGCGACGGGTCCGAGCTGTGCCGCGCGGCGGTCCGCTCAGGTG  
CAGCGCCAGGGGTGAAGTGCCTCTCGATCACCGCGGGCCGCCAATTCGCGCGGAGAGCGAC  
40 ATGCGCTTACGCTTGACCCCGCTGATCCGCGCCAGCTCGACCAGCTCTTCGCGGCCCTCGAT  
25 GGCTCAGGCTCGCGCCGCGGTACGTCTGCACCTGCTGACCCTGAACCCGCCCCCGGATGCC  
TCGSCGATCATCGCTCACAGCTACTACAGCCCGATGGCCTTGGCTCATGCCTTGGGCGCCAC  
GAGATCGCGCCTGTCTCGATCACCGTCGTACCGCGGGGTCGTGCGCGTCCGGAAGGAGCG  
45 ATTCGCGAGCCGCTGCAGGCGCTGATCGTGGGCGCGTGCCTGGTCATCCCGAGGAGTTCCC  
GGCTCAGCGTTCGGCTGCTGGACGTCAACGTCGACGATCCGGCACCGCGTCTGGCGGAGCGG  
30 CTCSTGGCCGAGCTCTCGGGCACGATCACATGGTGGCGCTGCGCGCGGCGAGCGCCTAGTG

5

10

15

20

25

30

35

40

45

50

55

GCCGATGTCGATCAAGTCGATGGCCTCGGTGTGGGGATCGCCAAGGTGCCCTTGCGCCGCGAG  
GGCCACTACCTGATTCTCGGCGGCCTGGGCGATATCGGCTACCACTGTGCCCCGCTATCTGGCC  
CAAACCTACCGCGCCAAGCTGACGCTGACCGCGCGTTTCGTCACTCCCGCCGCGCGCTCGTGG  
GAGCGAATGCTGCGCGAGGGAAACCTGGATTCCCGGCAGCGCACGCGCATCGAGCGCGTGTG  
5 TCGCTAGAGGCGTGCAGGGCCGAAGTCCAGACGGCTGCGGTGCACTTGGGCGATCGCCATCGC  
15 TTGGCCGATGTGTTCCGCGAAGCACGGGGCCGATTTCGGCGCCATCGCGGGCGTGATTCACTCG  
GCGGGGATTCCGGGACACGTCCACTCGATCGACGAGCTGGTGCGCGTCCGCGACGAAGCCCAA  
TTCACCGCGAAGGTTTCGAGGGCTGCACCACCTGGCCGAGGTCGTGATCCGCTGAACCTCGAC  
TTTTGTCTGCTGTTCTCCTCGCTCTCGACCGTCTCGGCGGGCTCGGCTACGGCGCCTATGCA  
20 GCGGCCAACGCCTACATGGACAGCTTCGCCCGCCGCCACGATCGGCCGGACGAATGTCGTTGG  
10 ATCGCGGTCAACTGGGACCGCTGGCTGTTTGAAGCCAAGACGTCGTCGGTCGGCGCCGAATTG  
GCGCGCTTGGCGATCGTGCCCGAGGACGCTCCGGCCCTGTTGCGCGGGTGCTAGAGCGACTT  
CCGCAATCGTTTCATCGTGTCACCGCCGACCTTCGGGCCCGCATCGACACTTGGATCCGGGAC  
25 AAGAACCGCGTCCCGCCCGCCGAGATCCGAGCGGTTCAACCGCGACCGGACCTGAGCCAGGCG  
15 TACGCCCCCGCGATCGGCCCGCTGGAGATTCAACTCTGCGGGCTGGTCTCCGCCTATTGCCGG  
TTCGACCGGATCGGGCGGGACGATTCTTCTTCAAATCGGCCTCAGCTCGTTGACTTGATC  
30 CAGCTCAGCTCGCGCATTACCGCATACCGGCAAGGATCTCAATACGACCCAACTGTTTACG  
TACCCACCGTGCGGCCTTGGCGCTCTCCTCGGCGGCAACCGGAGGGGCTCGCGGCGGAG  
GAGCCCGCCATGGAGAACCTGTGGCTGCAACGAAGCGATGCGACCTCGATGAG

20

**(2) peptide sequence**

Seq ID No 64 (&gt;pEPocos6\_ORF12.pep)

MTVEHETGFEIAVIGLACRVPGAADVAAFWRNLVEAKESVRFFEDHELRAAGVPPEEILRLPNY  
VKAKPLLADGEAFDADFFGFHPREAAAYLDPQVRLLECCWTALEDAGYDPAQYAYPIGLFAGV  
40 SSNLSFLFDRIDPRDSPLQKRYVAELNAASFATQIAYRLDLKGPASISITACSTSLVAIHLAA  
25 QSLIGGECHMALAGGATLEVPPKPGYLYREGYINSPDGHCRADFADAAGTIFGDGVGIVLLKR  
YRDALRDGDHVYAVIKGSAINSDGHRKVSYPGKSGQVAVIRAALAAQVEPQTI RFVEAHG  
45 TGTLAGDPIEVEALTEVF AEAGRCTCALGSVKTNIGHLDVAAGVAGFIKAVLALERRVLPPSL  
HFVRPNPAIDFNGPFYVCRQIERLTENGR LRAGVSSFGIGGTNAHVILEEAPAPPEARLPAGSP  
30 PGASPFLEPLSAKTPDALAGRCHDLADHLRAHPELLLADVALTLQMGRASFAYRHVVQAATAE

5

10

15

20

25

30

35

40

45

50

ELIRGLGAFRQESIRKRRNRVQWVLAGEAMSLDAGLRLYADWPVYRERVDVCLAIIVAKLRQID  
GRSFLHEWIERPREVPAEWSTALAFMFHCALAQALSQAGLHPQRMWSRGLGGQVGVVLAESLS  
LEQALALVLCQTPVPGDATPQRELVRTLLEGCRFRPPRFLISADSSGRPLDLAEFAHVDFWCG  
GQSASPNEAELRSWSDAAPELVTLAIGPSFLEAASGTVGLAIDPKRPMTCVQRTVAALWEWGC  
5 DVRWAAFTSSTGRRVPLPTYPFVRVIPTIGDPLRGAGAEDDLIAASASASAGSPPEPSANSAA  
15 ERPRAQSSIASATTPAPSHTSASVAVATILETVRAYFGFAAVRSTDAFFELGASSLDLVNLGQ  
LLSDRLGREVPTLLLYDHPTPDQLALALTSAAALSAEAPPLRGGHRASTSGTAASSAASTPTF  
PGDAHSQPSFVREQDIAIIGMAFRQPGADDLDAFWNNLVEGVESITFFSEDELLAAGVPREHL  
ASTRYVRKAGELTGMMDFEPEFFGYSAREAAVMDPQFRVFHECSWHALEHGGYDPTRCAASIG  
20 VYAGVTNHLPLWMLRTLPHLTEEBQFALLTDREFFAPLLSYKVGLRGPALSLQTACSTSLVA  
10 IGTACREL RAGACQMALAGGVTAIERCGYFHQEGYILSPDGHTRSFDAAGTVFGDGVGMV  
LLKPLAQALADGDTIHAVIKGIGINNDGARKVGFTAPSRAGQTEAIRAALRDAGVASNRVSYV  
EAHGTATRMGDPIEVEALTQAFRAEADGPLPPGSCLLGSVKSNGHLNAAACVAGLVKTVLAL  
25 QHRLPTSLFYQSPNPHIDFAASPFVRNGQTSDWVAPEGTRLLAGVSSFGLGGTNAHLIVEEA  
15 PKALPTTAAPLSTEPNDLDAGDADGLVLPISARTPTALAHATNLNHLERHPTIALADVALT  
LQLGRRQWPHRSLICRNRTEAIKLLRAVHSAEVPPAQAPVSDAPRCVFLFPQGGAQYPSMA  
RDLVRNCPDFALHLDPCLDQLAEILLPEDPRCILFGDGPADRLDQTAYTQPLLFSVSYALARWL  
30 GDFGIRPDAMIGHSLGEYVAACLAGLFSLSDALLLVSEGRMLGMSAARGAMLAVPLPEWELEE  
RLELLADDRISIAAVNTAESCVIAGPSEAIERCAQRWAAQGLTCTPLRTSHAFHSAMMEPIVE  
20 PFGHVLARVTFAPPRARWISNLDGKPIDSAAVMQPDYWVRHLRQPVRFHEGLSHLLAEDTHAW  
35 VEVGPGRTLSSFVRRHPAYRHQPIWNPMRHVESTGDVRRWRQALGELWRAGMPVAWERQRRG  
RHAGRRVPLPGYPFERRPFAARRFVELAQAPKAELVKNPDPARWLYRRVWRPAQAAAGGLAV  
QATVLVFGDGSEL CRAAVAQVQRQGLKCVSITAGRQFARESDMRFTLDPADPRQLDQLFAALD  
40 GSGSRPRYVLHLLTLNPPPDASAIIAHSYYSMPALAHALGAHEIAPVSI TVVTAGVVAVADEA  
25 IREPLQALIVGPCLVIPQEFPGLSWRLLDVNVDPPAPRLAERLVAELSGTDHMLRGGERLV  
ADVDQVDGLGVGIAKVPLRREGHYLLILGGLGDIGYHCARYLAQTYRAKLTLTARSSLPPRASW  
ERMLREGNLD SRQRTRIERVLSEACGAEVQTAAVDLGDRHRLADV FREARGRF GAIAGVIHS  
45 AGIPGHVHSIDELVRVRDEAQFTAKVRGLHHLAEVVDPLNLD FCLLFSSLSTVLGGLGYGAYA  
AAAYMDSFARRHDPDECRWIAVNDAWLFEAKTSSVGAELARLAIVPEDAPALFARVLERL  
30 PQSFIVSTADLRARIDTWIRDKNRWPPAETRAVQPRPDL SQAYAPPIGPLEIQLCGLVSAYCR

55

5

135

10

FDRIGRDDSFFEIGLSSFDLIQLSSRIHRITGKDLNTTQLFSYPTVRALALFLGGEPEGLAAE  
EPAMENLWLQRSDATLDE

pEP0cos6\_ORF13 sequences:

5

15

(1) nucleotide sequence

Seq ID No 65 (>pEP0cos6\_ORF13.seq)

20

10

25

15

30

35

40

25

45

30

50

55

ATGAAATACGAAACCACCGGATTGGAATTGGCCGTCATCGGTCTCGCTTGCCGCTTTCCAGGC  
TCACCCGATCCCGAACAGTTCCTGGTTCGAATCTGCGCGCAGGTTCGCTCCGGAATCCGCCATTTC  
AGCGATGCCGAGCTGAGCCACATCCCCGCATCCCTGCGTCACCATCCGCATTACGTCAAGGCC  
AAAGGCGCGCTGGACCACGCCGATTTCCAACAGCCTTCTTCGGCTACTCGCCCCAAAGAGGCC  
GAGGTGATGGACCCTCAATTCCGGCTGCTCCATGAGTGCTGCTGGGAGGCGCTGGAGTCAGGC  
GGCTATGCGCCGAGCCAATTGCGGGTTCGGATCGGCTTGTTCGCGGGCGCGCCTTCAACGAC  
GGATGGATCGCCGGTACCCCTCGACCGGCTGCGCACCGGCGTGGGTTTGAGCTCCCTGGAAACC  
GC3TTCCTTGACCCTGCGCGATTACCTGACCACCCAGATCTCCTATCGGCTCGATCTGCGGGGC  
CCCAGCCTGCTTGTCCAAACCGCCTGCTCGTTCGTCGCTGGTGGCGGTCCAGCTCGCCCCAGCAG  
GC3CTGATCTCCGGCGAATGCGCCCTGGCCTTGGCTGGCGGCGTGTGCGCGACCGATCCGCTG  
CATTTCGGGATACCTCTATGAACCCGGCAACATCTACGCGCGGACGGCGTCTGCCGACCGTTC  
GACGAGGCAGGCGCCGGTACGGTCTTCGGCGACGGGTGCGGCATGGTCCTGCTCAAGCGGCTG  
AGCGACGCCCAGCGCGACGGCGATACGATCTGGGCGGTCAATTCGCGGGGCGGGCGTGAACAAC  
GACGGGCACCAACAAGGTTGGCTACACGCTCCTGGCACGAGGGGCCAGGTGGCTTTGCTTAAA  
AGTGTATTATCGCGCGAGCCGGGTGACCCGGCGACGCTCGGCTACCTGGAGGCCCATGGCACC  
GGCACCGCGCTCGGCGATCCAATCGAGGTTCGAGGCGCTTACCCAGGCCTTCGCCAGCAAACGT  
CGCGGCACCTGCGGCTTGGGCTCGGTCAAGGGCAACCTGGGTACCTCAACACGGCGGCCGGC  
ATCGCTGGACTGATCAAGGTGGTGTGGCGCTGAAACATCGCGAAGTGCCACCCACCCTCAAT  
CTGCGCGCTCCCAATCCGAAAATCCGCTTCGACGAGACGCCGTTTTTCCCAGTCGTCGAGTTG  
CAACCCTGGCCAAGCGGGACCGGCCCCCTTGCGAGCCGCGTGAAGTCTCTTCGGCATCGGCGGT  
ACGAACGCCCACGTCATCCTCGAGGAGGCACCGCCGACGGCCAACCCGGCGCCACACGGCAGA  
TTCCGACTGTTGCCGCTTTCGGCCAAGACACCGGCTGCGCTCGAAGCGAAGCGCCGCGATCTG  
GCCGCTTCCTCGAACGCCACCCGGAGACCTCCTTGGCCGACCTCGCCTTTACCTGCAACGC



5

136

10

15

20

25

30

35

40

45

50

55

GGCCGCGAGGTCTTCAGTCACCGCGCCTGCCTCGCCGTGGAGACCTTAACGTCCGCGCGCACG  
CGGCTGAGCGGCGAGTCGTGAGCACTTGCGTGGTGGGCCCCGCGCCCAGCGCCATATTTCTG  
TTCCCTGGTCAAGGCAGCCAGCTCGCCGGGATGGGCCGCGGTCTGTATCACCATTTTCGAGCCG  
TTCCGCACGCGCGTCGATGCCTGTCTGCGCGAGCTGGAGCCAGGACTGCGGCAAGCGCTCAGC  
5 GCCCATTTCGATCCGAATCGCGGCGCGGACCCACCCGATTTCGACGACCTTCGTCCAACCCCTTG  
15 TTGTTCCCTCGTCGAGTACGGGGTGACCGAGTGGCTACGCTGCTTGGGTGTGCGGCCAACAAATG  
GTGTTGGGTACAGCTCTGGCGAGTATGCCCGAGCCTGCGTCGCGGGCGTTCTGTGCGCCGTCC  
GCGGCGGTCTCGCTGCTGGCCGAGCGCGAGCGGCTGCTGCGCGACCTGCCAGCCGGCGCCATG  
CTCGGCGTCCCGCTGGCCGCCGAGGCGCTCGAGGCGATGTTGCCCGACGCTCTCGATCTGGCG  
20 GCGATCAACGGCTGTGAGCTTTGCGCCGTGTCGGGCGCGTTCGCGGCGGTCCACGCCTTCAAG  
10 GCCCAACTGGAAGCCGCCGACATCAGCCCGCCTGTTGCACACCGATCGCGCCTTCCACTCG  
CGGCTGGTAGCACCGGTGCTTGACCGGTTCCAGGCAGCCGTTCAACACGTGGAGCTGCGGCGG  
CCGCAAGTACCTTACCTCTCGACCGTCAGCGGGCGATTGGAGGCGGATGGGCGGCGCAACCCG  
25 CACTACTGGGTGCGTCACCTGCGCGACACGGTGGGTTTGGTCCAGCCCTGGAGGCGCTGCCG  
15 CCGGTGGATTCCCTTCGTGTGCATCGAGGTGGGACCAGGCTCGGCCTTGAGCACCATGGCGCGC  
GAAACGTGGGTTCCCAGGCGCGACTGATTTCGTTGCTGCCGCGGCCGGAACGGGGCAAATC  
30 GAGCCCGGTCCGGTATTGGAACGACTGGCGGCGCTTTGGCGCAGCGGGTTGACATTGGATTGG  
TCTAAATTGACGGGCGGCGAAGAGGGTCATCGAATCCCTTGCCAGTCTACCCGTTTCAGCGC  
AGCCATCTGTGAGCTCCCTGGCGGCGGGCCACACGCCTTCGTGCGGCGCTGCAGTCGAATCA  
20 GGCGCCATCCCTTGCCGAGCGATCCGAGGGGAAAACGCTGAAACCCGGGATTGCCCGCTGCCA  
35 ACCGCCACGCTCGAGCCCAAGGCGGTGCTCCGGCCCCACTCGAGGCTACCGACGCCGAGGT  
ACTGCGGAGCGACTGGCCGAACCTTTGGCGCGAGTTGCTAGGGTTGACCTCGATTGGGCCCCGAC  
GACCATTTCTTCGACCTGGGCGGCCACTCGCTGACCGCCACGCGGCTGCGCGCCCTGATTAC  
40 CAGCGGTTTCGATGTGATCTCGGGCTCGACGAAATCTTCGCTCATTCGCGTCTCTCCCAGCTG  
25 GCCGCCCCGTATCGAGGCGGCGGCCAAGAGCCGATTTTCTCCATTCCCAGCGCGCCGACCAG  
GAC3ACTATCCCTTGTCATCCGCCCAGCAGCGGATTCACAGCATCGTCACGAGGGCCGAGGTC  
GGCACTGCTTATAATTTCCGATCGTCCTCGAGCTGCAGGGCGCTCTGGATCGAGTGCGATTTC  
45 GAGCGGACGTTGCGGGCATTTGTTCCGGCGTCATGAGGGGTTCCGCACCCGCTTTGTGATGCGC  
GATGGCGGGCCGCGCCAGCGCATTGTACCGGACGTGGCGTTTCGCCTGCCGCTACCCAGGTC  
30 GAGCCAGAGCAGGTTCCCGGGCGCATCGAGGCCTTCATCCGTCCCTTCGATTGGAACGCGCG

5

137

10

15

20

25

30

35

40

45

50

55

CCGCTGTTCCGCGCGGAGCTGTTGCAGTTGGCCGAGCAGCGCCATCTGCTACTTTTCGACATG  
CACAACCTTAATTGCCGACGGTATCTCGCTCAACCTGTTTCGTGCGCGATTTCGCGGCCCTGTAC  
CATGGTCGTCCGCTGGCGCCGCTGAAACTCCGCTATCGCGACTATGCCGTTTGGCAAGAGGCG  
CGGCTGGCCTCCGATGACCTGCGCA3CCAGCGCGAATGGTGGCACC GGCGGCTTTCGCCGCCG  
5 GTCGCCACGCTGGCGCTCCCTCCCGATTTCCCGCGTCCGGCGGTGCGCCGCTACAAGGGCCGT  
AATGTGGTGTTCACCTGGACCGGGAGATCCGCGACCGCCTGGTGGCCCTGGCTCGAACCAG  
GGGTCACCATGAACGTGATGATGCTGGCGCTCTGGGCTGCGCTGCTGCATCGCGAAACCGGC  
CAATCGGAGCTGGTGGTCGGATCGCTGCTCGGCGGGCGGCCGACAGCGAGCTGCATCCCGTG  
ATCGGGCTCTTCACCAACTTTTGCCCTTTCGGGTTGGCGGTGAGGGATCGACCCGCTTCGAT  
10 CGCTTCCTTGCCGCTTGCCACCAGGTGTTTCTCGAAGCCTATCAGCGCCAGGACTATCCGTTT  
CACTTGTTAGTCCAGGAACTCGTGCCGGTCAGGGACCCGTCGCGGTGCGCGCTGTTCCAGACC  
TCGCTCGTCTACCACAACGAAATTGACGGCAAGACCAAGCTGGAATTGGAAGGGCTGAAAGTC  
GAAGTGGTTCCCTTCGAAAAGGGTGTGGCGAGGCTGGATTGAAGCTGGATGTGACACCTTTT  
TCCGACCGACTCGAATGTGTTTTGCAATACGACTTGGATCTGTTCTGCGAGGAGACGATGCGC  
15 GGCTGATCGCGCGGTTCCAGGCGTTGGTGGCGGGGCTTGTGCGCGATCCGGCGCAATCGCTC  
GCCGCCGCGAGCGTTTCCGGGAAGCGGGCGCTGCGCGCGGGCGTGGCCACGGCAAGCGAATCG  
TCGCCGCGAGTCACTGCCGCCGCAACCATCGACGGCGTACGCCACTCCCTCACCAGTACACCG  
TCGCCGGTAGTCCTGACGGGACCCGCCGACCTGCCCGGATCTTGGCGGCCTACGTGGGGCAG  
AACCCCATCCGTTCCGCGATCCATCGGGGTCTCATTTTGGAGGCGCGCTGGGGTTGCGAGCG  
20 CTGCGGTGCGCGCTGGACGCGAGTGTCTGGAGAACACACCCATTGGCGCAGCGTGGTGGCGG  
GATCGCGCGCGGCGGTGGATAAGTTGGAATTGACCAGCCTGGTGGCGCTCGACGACCTGCGC  
GGGTGGTCAATCCTCAGGCGAATGCCTTACCCCTGGCTTGGCGCGATCTGGCGATGCCGTTT  
GGGGAGGGGCGTCCCTGTGGCGACTCCGCCTGGCGTGGTGGCTCCATCGCGCTGGTTGCTA  
40 TTGCTGACGGTTCATCCATTGATCGGCGACAACGGCACGGTCGACCTCTTTCTGGCGGCACTC  
25 GCCGATCACCTGCGCCGCGCGTCCGCTTTTCCCGTAGCACCGCTCGATGAGGCCGAGCTGGAG  
GCCGAGCTGAAGTGGGGAGAGGAAGGGGAGGGCCTCGGGCTGACCGGATCGCGCCGGTCTCTG  
GGCCAATTGCGCGAAAGTCGGCTGAGTCTGTGGCCAGATGTGGCTGGACGAGGTCTGTGCG  
45 CGCCACGACCTCACCCCGCTAGAGGTCTTGGCGGCCCGGCTCCTCGATTGGACACGAAGCCAC  
GGTCACGGGTGATCGCTTTGTGGACGCCGCTGCCCGAGGACCATCCGCTTCGCGATGAAGGC  
30 CGCTGCCTCCAGGTTGCGCTGCTGGAGGGGCGCCGTCGCGAGCGAGGAGCGGGCGATCCAAGC

5

10

TGGCTCGAGCAAATCGCCTTGAGACGGGGTACCCCTGCAACGGAGGTCGTTTGCCTACTCCG  
ACCCAACGGGCAGCCATCGACCTCGCGCTGGCCTGGCTGCCGAGCCGCCTCTTCACGGTTTG  
GTCCGAACCGTTTACGCCGTGGCCGGAATCTCCATTGGTCTGTCCGTTTCCCCTCAATCTCGCG  
TTCCGGCCAAGCCATCCAATTGCCTACGCGCTCAAGCACGAGGCCACGCTCGCGGTACGGCA  
5 CGGGCGCGCATCTGATGCGTTTCTCTGACGGCTTGGGCCCGGAAAGC

15

(2) peptide sequence

Seq ID No 66 (>PEPocos6\_ORF13.pep)

20

10

25

15

30

35

40

25

45

30

50

55

MKYETTGLELAVIGLACRFPSPDPEQFWSNLRAGRSGIRHFSDAELSHIPASLRHHPHYVKA  
KGALDHADFEPAFFGYSPKEAEVMDPQFRLLECCWEALES GGYPQS QFAGRIGLF AAAAFND  
GWIAGTLDRLRTGVGLSSLETAFLTLRDYLTQISYRLDLRGPSLLVQTACSSSLVAVQLAQQ  
ALISGECALALAGGVCATDPLHSGYLYEPGNIYARDGVC RPFDEAGAGTVFGDGCGMVLLKRL  
SDAQRDGTIWAVIRGAGVNNDGHHKVGYTAPGTRGQVALLKSVYRASRVDPATLGYLEAHGT  
GTALGDPIEVEALTQAFASKRRGTCGLGSVKGNLGHNLTAAGIAGLIKVV LALKHREVPPTLN  
LRRPNPKIRFDETPFFPVVELQPWPSGTGPLRAGVSSFGIGGTNAHVILEEAPPTANPAPHCR  
FRLPLSAKTPAALEAKRRDLAGFLERHPETSLADLAFTLQRGREVFSHRACLAVETLTSART  
RLSGESSSTCVVG PAPS AIFLFPQGSQLAGMGRGLYHHFEPFRTAVDA CLRELEPGLRQALS  
AHFDPNRGADPPDSTTFVQPLLFLVEYGVTEWLRCLGVRPTMVLGHSSGEYAAACVAGVLSPS  
AAVSLLAERERLLRDL PAGAM LGVPLAAEAL EAMLPDALDLAAINGCQLCAVSGPVAAVHAFK  
AQLEAAGHHARLLHTDRAFHSRLVAPVLD R FQA AVQHVELRRPQVPYLS TVSGRLEADGPANP  
HYWVRHLRDTVRFGPALEALPPVDSFVCIEVGPGSALSTMARETLGSQARLISLLPRPRTGQI  
EPGPVFERLAALWRSGLTLDWSKLTGGEEGHRIPLPVYPFORSHLSSSLAAGHTPSSRPAVES  
GAILAERSAGENAETRDCPLPTATLEPKAVAPAPLEATDAAGTRERLAE LWRELLGLTSIGPD  
DHFFDLGGHSLTATRLRALIHQRFVDVLDGLDEIFAHSRLSQLAARIEAAAKSRFSSIPSAPDQ  
DDYPLSSAQQRISIVTRA EVGTAYNFP I VLELQ GALDRVRFEATFAALFRRHEGFRTFRVMR  
DGGPRQRIVPDVAFRLPLTQVEPEQVPGRIEAFIRPFDLERAPLFR AELLQLAEQRHLLLFDM  
HNL IADGISLNL FVADFAALYHGRPLAPLKLRYRDYAVWQEARLASDD LRSQREWWHRRLSPP  
VATLALPPDFPRPAVRRYKGRNVVFHLDREIRDRLVALARTQGV TMNMMLALWAALLHRETG  
QSELVVGSLLGGRPHSELHPVIGLFTNFLPLRLAVEGSTRFDRFLAACHQVFLEAYQRQDYPF  
HLLVQELVPVRDPSRSPLFQTSLVYHNEIDGKTKLELEGLKVEVVPFEKGVARLDLKDVTPTF

5

139

10

5

15

20

10 pEPocos6\_ORF13.1 sequences:

## (1) nucleotide sequence

25

Seq ID No 67 (&gt;pEPocos6\_ORF13.1.seq)

15

30

20

35

## (2) peptide sequence

Seq ID No 68 (&gt;pEPocos6\_ORF13.1.pep)

40

25

45

50

55

## pEPOcos6\_ORF14 sequences:

## (1) nucleotide sequence

Seq ID No 69 (&gt;pEPOcos6\_ORF14.seq)

5 ATGGTGACGCGTCCGACGTCCGACGGCATCGAGGACGAGCTCGCGCCGTTCCTCCCGGTCCTG  
15 CGCGGCTGGCTCATCGAGGGCGAGCTCGGCCGCGCGGGATGGGGCGGGTGTTCGGGCGCGG  
CACCCGAAGACGCGGGCGCGGGCGGCGATCAAGGTGCTGCTCGCGACTACGCCCGCGGCGCG  
GACGTGCTGGCCCGCTTCGGGCAGGAGGCGATCGCCGTCAACATCATCAACCACCCGGGAATC  
25 GTCCGCGTCTTCGACTCCGGCGAGCTCGAGGACGGCTCGCCCTACATCGTGATGGAGTACCTG  
10 GACGGCCGGGGGCTGCGCGACTGGGTGCAGGCCGTGCCGCCCGCGGAGCGCCGCGGACGGTC  
GTGCGGCTCGGCTACCAGATCGCCTCGGCCATGGCCGCGGCGCACGCGTCCAAGGTCTCCAC  
CGCGATCTGAAGCCGGAGAACATCATGGTGGTTCGAGGACGAGCTCGCGCCCGGGGCGAGCCGC  
25 GTCAAGATCCTCGATTTCGGCATCGCAAGGTCTCTGGGGAGGTCTGCCCGAGGTGCTGGAG  
CTCGAGGGGAGAGGCTCCCTCGCGCCCGCGTCCGCGTCCACGATCCGCACCGAGCTCTCGACG  
15 CGGCCGCGCGCGACGGTGGGCGCCACGACCGGCCAGAGAGCCCGCTGGGCGCGAGCGCCACG  
CCAGAGAGCGCCCTGGGCGCGAGCGCCACGCCAGAGAGCGCCCTGGGCGCGAGCGCCACGCCA  
30 GAGAGCGAGGCCCACGAGGAAGACGCGCTCCGGAGCCTCCCCGTCTGACCAGCGGCAGGCCC  
GCGATCCACCCCGCGCCGTCGAGATCCCGCCGAGGCGGTCTCTCCGCGGCGTCCGCGGGG  
TCGCGCGCGTTCGATCGAGCCAGGCGCGCCCGCGCCGAGAGCGAGGGCGCGGGACAGCCACG  
20 ATCCCGTTCACGCAAGAGGGCGTGTGGGGCCTCGGGACGAGGAGCTACATGGCGCCGGAGCAG  
35 GAGCGCCACTCCGGGAGCGTGGACGTGAAGGCGGATGTCTACTCGCTCGGCGTCATCCTCTAT  
GAGCTGCTCGAGGGGCGGACGCCCCGACGCGCCGAGCGCCGCGTGGCCGCCCCCGATGAGCGCC  
GCCACGCCGCCCGATCTCGTCGCCCTCGTCCACCGGGTCTGGCGTTTCGATCCCGATGCGCGG  
40 CCGCGCATGGCGGAGGTGGCGAGCGCGCTTCACCGGCTCGGCCGGGCGAAGAAGGAGCTCGAC  
25 GAGGCGCTCTCGAGGTGGGTCTCGGCGGAGGGGCGCCGGGGCTCTTGCCGTGCGGCTATGCT  
CTTCTCGAACTGGTCCTCCTGGGCCCTGGGAACCTATACGATTCTTTCCAGCCTGTAAGTGCA  
TTTCTCTTTCAATATCGTCCTCTCTTCATATACGAGGTGAGTTCTCTGAGGTCTCTCTATAAG  
45 TCTGGGGTGTCTATTTCGGCCTCTTACTTGTTACTTCGCCTCTTAGGAGTTTTCTCTAATT  
TTGCCCTCTTACATTCCCGTATTCACTTAAGTGGGCCCTATCTCATTCGC

30

50

55

## (2) peptide sequence

Seq ID No 70 (&gt;pEP0cos6\_ORF14.pep)

MVTRPTSDGIEDELAPFPPVLRGWLIEGELGRGGMGRVFRARHPKTRARAAIKVLLGDYARRP  
DVVARFRQEAI AVNI INHPGIVRVFDSGELEDGSPYIVMEYLDGRGLRDWVQAVPPAERPRQV  
VRLGYQIASAMAAAHASKVVHRDLKPENIMVVEDELAPGGSRVKILDFGI AKVLWGGLPEVLE  
LEGRGSLAPASASTIRTELSTRPAPT VGATTGPESPLGASATPESALGASATPESALGASATP  
ESEAHEEDALRSLPVVTSGRPAIHPAPVEIPPEAVSSAASRGSRASIEPGAPAPQSEGAGQPT  
MPFTQEGVWGLGTRSYMAPEQERHSGSVDVKADVSLGVILYELLEGRTPDAPSAAWPPPM  
ATPDLVALVHRVLA FPDARPRMAEVASALHRLGRAKKELDEALSRWVVGGA PGLLP  
CGYA  
LLELVLLGPGNLYDSFQPVSAFF FQYRPLFIYEVSSLRSSYKSGVSY SASYLLLRLLRS  
FSLI  
LPSYIPVFILTGPYLIR,

or DNA sequences complementary to said open reading frames,

(b) DNA-sequences which hybridise under stringent conditions to regions of DNA sequences according to (a) encoding proteins or to fragments of said DNA sequences,

(c) DNA-sequences which hybridise to the DNA-sequences according to (a) and (b) because of a degeneration of the genetic code,

(d) allele variations and mutants resulting by substitution, insertion or deletion of nucleotides or inversion of nucleotide segments of DNA-sequences according to (a) to (c), wherein the variations and mutants offer isofunctional expression products.

10. DNA sequence according to any of claims 1 to 5, wherein the DNA is selected from the group consisting of

5

142

(a) the following DNA Sequence:

10

Seq ID No 71 (>Contig43)

15

20

25

15

30

35

40

25

SEQ ID No 72 (>Contig44)

45

30

50

55

CGGGTATTTGTGATATGTGGGCNGTAGTCGTATGCTTCATTAAGTACATC  
CGTCCGTINGTAGAGAGTGA CTCTGTGCGCAGCGATAATAGACACGCTTGTG  
ATGCTATAGGGAACATAGAGTCNTAGTAGATGATACGACGAGATATTNGT  
ATAGAGCGTATAGACCGACGTGTGAGCGTCATAAGTGTTGTGTGTCATGA  
GTGTGCTCAGAGGACGTGCAGACATTATATGAGCAGATCATGAGAGAGAA  
TCAATGCTGCAAGNTATTCGTGCAATCTACATTATATCGAATCGTGTATG  
TGCCTTTGTGCGCAGCGGATNCGATGAGATACCGAAAGGGTATGTATCTA  
TNTTCGTGACGCTCGATNAGAGCAAATCCGCTACCGTGGAGATATCGTGT  
ATCGACTCCATCAGCATCAGTATCATGATACGTCAAACGAGTACACTCAT  
TATTGATAACACACGTANGTGTGCATGCACAGTTATCGAGTGTATTGTGT  
GCATGAGAGGTATAGGATNTATAGGCGAGCATATATATCTATATATATAG  
GTTAAGAGTAGAANACTATGAAGATGCAGGAAGTAGTATCTCGCGGACAA  
ACGNGTAGCTAGCGGGGTGAAGTATTATCGACAGTGTATAACGACTCA  
ACAGGGNTACGAGGTACATTGTATTTACAGTGTTGGAAGGATTGCGCGA  
GGAAGGTAGTGGTACCGTGTGAGCTACGATGCTCGGGATAATGGTGATT  
AGATAGAACCCTTAGCGTTGCTAGATGAGTGAGTGGTGGTATGAGTAGAGT  
TTTGTCTAGCTTTGTGTCCAGCGAGGATTCGTTTCAGTCTGAAGGGTAA  
GAGTACGTCCATCGCACACCCGACCGTTTTGAGGAGTTCTCGGTGCGTGG  
TCAGTGGGGTTTGGAGAAGACAGAGTTGATTCATAGGGTTATCAAACGAG  
TTATGTGGATAGATGGTAGTGACCCCATTTGAGTGAGAGTGTGGCGTTA  
ACANCAGCAGGATNTAT

TAGTCTTTGACACCATGGGAGCTGCTACCGATGTTGCCGAGCACGATCG  
CGCTGGCGCCGACGAGCGACTGCAAGCCGCGCCGCGCCATTTACGCCTGA  
CGAGCGAGGTGGGCGAAGTGCTGGTGCGCGCCGTGCGTGTGAGCGCGCC  
CAGTCCCGCGTTGCGCCGTGCGCCGAGCAGTAGCGCGCCGTGGAAGACGA

5

143

10

15

20

25

30

35

40

45

50

55

TCACCGCGATCGAGGTCAGCGTCGTGGGGGCGAGGCCGAGGAGCGCGAGG  
ATGCCGAGCACGACGCCGGCCGCGCCGTAGACCAGCTTGGCGCCCATGCC  
GCCGCCGAGCTCGGTGCGCGTGTCCCACTCGACGGGCGGCGCGGTGCGGC  
TCAACGCGCCGAAGCGCGAGGCGATCGCGCCGCCCTGCGCGATCAGCGCG  
5 GCGCCGAATACGATCGTGGCGATCTGAGGTGAGCTCTACTGGCATGATCC  
15 CCGTCAGCCCCGAGGATGGTGAGGACAATCGTCGCGGCGCCGCACAGTACC  
TCGCACGAGCGAGCCTCCGAGCACGACCTTCGGCGTCGTCTCGTCCTTTG  
GTCTGCGTCGCGCGCCCGAGTGCGGCGTTATGTGGCTCTCCGGCTGTGCA  
AACCGTTCACGTTCTTCCGGTCTTGAGTCAGCATCGGCATGATTCCCCC  
20 GTCTGCGGTGAGGCCCTTGTGCGGCTCACGCGGCTCCGACTTGACAGTG  
10 CTGTGCCGGGTCTCTCGCTCAGGAGGCGCCTCTCTTGGTGGTGCTTGCG  
TCCTGGTCCGTTTGCCCGCTGTGCGGTAGGTTTCTTGAACCAGGTGACC  
TTCAGGAGCCCTTGATGCGCTCCATCGTGTCTATGTCGATCCTTCTCT  
25 GACTTGTATGGGTCTCGAACCAACTACGCTTGATCAGGCCTTCGAAGGGT  
15 CCTTTGGGAGATCGACTCTGGATCCATACCGGGAGCCCTGTTCTGCCGC  
TCTCTTAAGTTTCCCTTCTGTATCCGTGTGACCGGAAACGCTTTATCT  
30 CTAATGCGCTCTAATTGCGTCTCTGCCACACGTGCGCTTCACTCTGGATC  
TACTTCTTCTCCCTAGTCTTCTACCTCCGTACCCTTATTGTTGGTTCTA  
TTTATTTCTTTTCGCTTCACCTCGCGTCATTGTGCGCTAGTGTTCCCTCCC  
20 TCATATCGCCTTTGGTCTCCCTCGAGCGTACAGTCCTCTCTTTCAGATG  
35 CTTTCCGGCTCCTCTTCTGCTGGCCCCCTTATCCTTTCTAATACTTC

SEQ ID No 73 (&gt;Contig48)

ATGCGCCAGGAACACCCCGGTGCGGCTGCCGTGAGGGACTGGGGTGCG  
40 ATGCCGGCGTCTCGAGCCCTTCCCAGGTGACCTCCAGCAGCAGGCGTTG  
25 CTGAGGATCGAGCGACCGCGCCTCCCGAGGCGAGGTGCCAAAGAACGCGG  
CGTCGAAGCCGTCCACCGCCTCGGTGAGCAGTCCGGCCCAGCGCGGCACC  
45 TCCTCGCTGGGATGGACGCCGACCAGCGCCAGCGCCGGTCGAGCGGCTG  
GACCGCGTCTCGGCCTGAGTCGAGCAGCTCCCAGAATGCCTCCGAGTGT  
30 CCGCTCCGCCGGGAAGCGGCAGCCAATGCCTACGATGGCGATCGGCTCG



5

10

15

20

25

GTCCGCTCTTGCTCCAAAGACGCGTTCTTTTCGCAAGCTTGTCCATGAG  
CAGAAGGGCATGCTCAAGCTTCCCGGCATTTCGTGGTCGCCATACTCCCTC  
GGTCCCTTACTCACCAACGATCTGCGCGAGCTGCGCCAGCTTTTCGGCGA  
GCAACGCGTCCTTCTGCTCGTCCGTCATGCCCCGAGAGCCTCGAGATCT  
5 GCGGCATCGTTCTCGAAGCTCTTCTCCCGCTCGGTGGCCGAGCGTGGGT  
CGCGCCGGCATTGGAACAGAAATGTCTAGCAAGCTCCCGCTCAGAGCTG  
CTACGTTAGGGTAGGTCCATAGCAGGGTCGCCGGCACGGTGATGCCGAGC  
GCGGCCTCGATGCGGTTGCGGAGCTCCAGGCCTATCAGCGAGTCCATGCC  
GAGATTGCTGAACGGCACGTGCCGCTCGATCCTCTCCGGCGGAAGGCGCA  
10 GCCCCCGCCCCAACAGCTCGCTCAAGTGCTTCTCCAGAATCAACTGACGA  
TCTTCGGGCTTGGCGCTCTGCAGCGCCTCGCGCAGGTTGACGCGTTCGA  
CGCGCCTCGGTGCGGCGCGTACGCTCCTTCAGCAGCTCCGCCCACAGCG  
CCAATCGGGCCGCGTTGGGATAGAACTC

15 SEQ ID No 74 (>Contig49)

30

35

40

45

50

ACCACCGCTTCACTCAGTATGTACTTTGTTATACTCGTCTTAGTACAATG  
ATATAATACTCATGTGTATTCTTAATCTCGGGGAGANAAAATTGGAATAC  
TGGACACCGTTGCCGCATGCNCACTCTAGAGATCCCCCTGCGACGGTATC  
CCACGGCACCGGTATGGCCGGCGCGCTCCGGGGGTCAACGCCCCGTGG  
20 TTGCCTTCACGACAACCGCGTCCGGCGGGCGCCGTTTCGATGCCGCGGG  
CCCGCGCGGGCGGCGCTTATCCTGTGGAGCATCTGGAGGGCGCTCACG  
CACCTGTCACTCTAGTTCTGGCCCGCCCGAAGGAGTCCGGGAGGCCGAA  
GTTGAACCCGATGTAGAGCGCGATGAACGACGGGAGCACGCGCGCGGGGA  
TGTGCAGCGCGGCGCCGATCGGCGTCCGGAACAGGACGAGCTCGCCCGGC  
25 ATGCCGGGCACGACATACCCGAGCAGAAACAGATCGGCACCACGAGCGT  
GAGCTCGAGCAGCGATATTTTCATGACCGACCGCGGGCGGCGGCGCCCG  
CCATGACGAACACGAGATCAACGTGCCGTTGACGTTGAGCCAACCGCCG  
AGTCCCACCACGAAGAGCCTGAGCTCCTGCGGCACCGCCGGATAACATTT  
GCGGACGAGGTGCAGGTTGAGCGGCGTCCGAGCGCCTCGCTGCACGAGG  
30 CCCACAGCAGCGGATAGACCTTGAGCCAGTAGTTGACGAAATAGTCGCGC

55

5

145

10

15

20

25

AGCGAGAACTCCGGGGCGGCGGCCTTCATCCGCAGCAGGCTCGCCGCATG  
GAAGACGAGGCAGGCAGCGCCGACGACCCCGACACGAGCAGATAGGAGA  
GCATGAGGTCTCCGGCGCCGAGCGAGGGCGCCCCCGCGAGGCGCGCGCG  
AGGTCCCCCCGTGCACCTGCGCGGCGAGCTGCGGGGAGCCCGCGGAG  
5 ATAGGCGCCGAGCCGAACATGAAGAGCGGGACCAGGCACTGCACGGCGC  
CTCCCGCGCGCTCCAGCGCGTCCGCGCGCGCTCCAGCGCGGGGCGACC  
CGCGGCGCGGTACGGCCGCGAACGACGTACGATGCCGCGGTAAAGGGC  
GAGGAAGCACGGGCTCGAGATGACCAGGCCCGAGGCGCTATAGAGGGTGC  
GCGCGGCTCGAACGGCGCGCCCGTGTGTGGCTGGGGAGCAGCGGCAGC  
10 CCGAACACGAGCCATGTGACGACGACCCCGAACAGGCACGCTGCCAGACG  
CTTGAGGGCGAGCCAGCCCATGATGACGCGAGCAGCCGCCCGGGCGCC  
CTTGCCGGTGCAGGCTCACGAAGGTGGCACGAGGACGACGAAGATGACG  
ACCGGCGCCAGCGTGGTGTACCAATGCAGGAGACCGTCCATGGCGCGGGT  
CGACCACGCGTGACGCTGGTCTCTGTCTGACTCGATCATGGCCCATT  
15 CGCCTAAAATAATGATCCGTTCTCAAATGGGTCAAAAAAAGTTCCCTT  
AAGACTGTTTTACTCCGGAATATTAATATATTTCTGAGTGTGAGGTGATG  
TTAATCACACATTCTGATATTCTCAAGGGGAATCCGTGTCATTGTGAATA  
CTTCTCTCTCTACAAGAGAGGTTATATATGGTCTCGAATATCTCGTCCGC  
TCTTATATATATTCTCTGTGATAATATATATCGAGTGTGGGTACTCAGC  
20 TCTCTTGGTGTAATCTATAACTCGGCATCTCTCATAATACCTTATATATA  
CACACTCTCTCGGTCAATCTCGCATAATAGATATATTTATATGTTCCG  
CGTTTTATCCGAGTGGGATACACTTTTCTATATTTCTTTGGTGTGACG  
CGTGGCGTCGAGCCTTATTATTGATTGGTAGTCACGATATTCTCTAGAT  
GACATCATACAGATGCTCATAACTCGATAAACACAGGTGCTACACGACGA  
40 GACTCTCACTCTCACTCTT  
25

45

SEQ ID No 75 (>Contig50  
TCCCCAGTTTTCTCCTCTCTACGCTCAGCAGGAAAAANATAAT  
GGAGAAATCGTTGCGCTCTAGCAGCATCTATAGGATCCCCGCTGCTCTTCT  
30 TCATGCACCTCGTGGAGCAGAAGTTCATCAACGCCTTCGCGATCATCGTG

50

55

5

10

15

20

25

30

35

40

45

50

55

GCGGTGAGCTTCCTGGCGTTGCTCCTGTGCTCGTTCGTGCGCGACGTTCGC  
GACGCGGAACACGTTCCCGCCCGCGCCTTTGCCGGCGCTGAGCCCGCCGG  
CGCCGGCGCTGATCCCGCCGGTGCCGGGCGGATCCGTTGGGCGGTGCGCG  
GAGCCGCTGTGCTCGGCCGGTGATCGGTTGTGCGGGCGCCGTGCCTCGG  
5 GCTTACTACCCCTCTCGCGGGTGCGGATATGGCCGTGGATGAGGGAGGC  
GATGAAAATCGTGATCGCCACGTGCGCGTTGTTCTAGATCGTCCCAGGCT  
15 GACCGTCGGGAGCGCCAGCACGAGATGAAGAGCCACACCGCGAGGACCG  
TGTTGAGGTACCGCACCAGGGGCGAGCATGGCGGTGATCGCGAAGATC  
ATGCAGAGCAGCCCGAGCACCCATGTGTTGCTCCGCTGCGCGTGGCTGTG  
20 CGGCCAGATGACGGCCGAGATGAGGAGCCAGAACCCGAGGACGACGTTCA  
10 CGATSCGCGCCATGAGATTGCCAGCTCGAACCATGCTCCCTCCACCTCC  
GATCATGGGACCGATCGGGTCGCCACGGATCGATAACGGGCGTCAGGAGA  
CGTCAATCGGCGAGCTCGTGAGCCATGGCGACAGCCCGCCGACCGCGCC  
25 GCGGGTCTCTGGCTGCTGGTCGCCGTGCCGGCGCGGCGATGGGCCTG  
15 CCTCCGGTCGGGCCGCGCGGGCGCGCGCGCTCCGCGGAAGCTGGAGAC  
GCGGGACATCGCTCCCTCGCCCCGCGCTCGGACGAGCGGCGCGAGCCACT  
TCTCGACGGCCGAGCGGGCACTAAGCTTCCGTCATGAGGCTCGGCGCACG  
30 GCTCACCACGCACACGTTCTCGGCCGGCGCCGCCGCGCATCAGCTTCGTG  
TCCAGCCGATCCCGGGCTCGGACCAGCTGTTGTCATTCCGATCCAGTAC  
20 CTGCTCGGGCGTCGCTCGCGAAGGAGCGAGGCGCGCGCTCTCGAAGGC  
35 GCGGTGGTCCCAGGTCCACCAGCTCATCTGGGGCGGCGGCGCTTCGCC  
TCATGCTCGGCTTGACCCTAGGGCTGATCCCGCTGGCCGGCGCGTTCACG  
AACGCGATGACGGCGTTCTCACGACCGAATATCTCGGGTACTACGTGGA  
40 TAGAGCCCTCGACAACCCGGACAATCCGCCTCCGCCCTGTGATCCAGG  
25 ATGTCTTGAGCCCATCACCTCCTTCTTACCGGGCGAGCGCGGTAGGCG  
AGC3GTCCCTGGGTGAGCCCAACCTGCGGCTCTAGGAGCCGAAGGGCGA  
GCTCCTCGGGAGCGGCGCGCGTCAACCAGATTGCGCGGCGCTTGCGG  
45 CCG3AGCGTATCGCGACCGCCGCCACCGCCGCCACGGCGAGCACGGTGAC  
CG33GCGGCCCGCGGATGGCGACGCTCCGGGCCGTGTGCTCGGCCTCGC  
30 CCTTGATGCCGCCCCACCTCGCCTTGACCGTGACCAGGTGCGCGCGGAGC

5

147

10

15

20

25

30

35

40

CGCTCCTTGCTGTGCTCGATCTGCTCGGCAGGCCCGCGGGCGCCACGAA  
ACCTGCGCCGGCGCTGACGTGACCTCGCTCGCTGAGGGGACTGGCGACCC  
TCTCCGTGTGGAACCGGATTGCGCTGGATGGATCCATATGTCCCGCGCTG  
CAATCGTTGCTCCCCGCCGGCACATCGGAGTGCTCGCCGGATCGCGCGGC  
AGCGCCGACGCCGTACTTCCATAGGATAGCCCACCCCATCGGACAAGCCG  
GCTCCTGACGGCGGGCACCGAATGTTCCGCCAGACGGGCACAAGGCGCACG  
CCGCGGACGGATCGGCCGCACTGGCACTCCAGAGCGCATCGACGGATGGC  
CGACGGATGTGCAATGAGGCGCCCGCACGAAGCGAATTGTCCCGAATACA  
GCGAAGAAATCTATAGCGATGCGAGCAGAAGGATATGTCTATGGGGGGCA  
GTGAGAACTGGGGACAGTCAACGCACATATTCTCTCCAANTGCTAACGA  
CAGCGTGCGCAGAGGAAGTATCCTACTAGTGTAAGAGGGACATTTCGATGC  
GACCGCATAAACATTTCAGTCTACAACCGGTGAGAGGATGGAACACCCCGC  
CCCTCTGAAGGCTAGACAACCATGAATATGTGCAGAGGAAACACAGAATT  
CCAAAGGTGAGAACATATGTAGGATCGCGCCACCCGAGATTGAGTGAAGA  
TATACATATATACTTATATGGATCTACAACATGGCGAACCGAACGTAGCA  
NAATAGTAGATATAATTGTAATACTGAGCTACCGACAGAAAGATACACAC  
GAGTGTAACACATCACACGCAGAGTGGTACCAAATTCACACCATGCGAG  
CCACAATGTGACACGGAGGAGCACAGCATGGGCGCCACTATGGAGGAGAA  
ACTACTGCAACCCACATCTGATGGACTGACCGCACGGACGGGACGTGTCT  
ATACATACAGATACATCNGATGGAGGAAGATGCATGTGCGATGATATCAT  
CGTCGCAAACTCATATGTGGAAGAAGATATGNGTCAACTCAGCACTACTC  
ACACGATACGTGAACAGGAGTGACTAGGACATCNCATGGTGTGTCGGCGC  
GTGCACGTGATATCAAACCTCTGATCAACCACACACTATATAAGGAGTA  
TCGAGCGGCGATGGAACACCCCTCACAGCATACGTATATGCACAACGTC  
TGAACACTCTNGAGACACAGTGAAGG

45

50

55

SEQ ID No 76 (>Contig51)  
GATCCAGTTACGCCCCGCCCTCGGTACGCCGGGGTTTTTCGGCGTCGA  
CCGCGGACGGTCGGGGCGACAACCGGGGGTTGTGCTCAGCGGTTGCGGTG  
GATGTGCGCGACGAAGTTGTCCGCCGCCGAGTCGTCTGCGTGCCGCCGC

5

10

15

20

25

30

35

40

45

50

55

GCTGCCGTACGAGCAGCCCGAGGCACCGCGCGAGCTCCTGCCGCTTCCGA  
TGGCGGGCGGCTGCTCGGCCTCGTGCTCCGCGATTCCGGCGCTGGCGCCT  
GCTCTCCTCGGCGCGTTCTCTGCGTCCGCGCGAACAGCACCTCCTGCACCT  
CCGCGAGTTCTCGGCGGTAGCGACCCGCACGTGATCCGCACGTCTCTCCG  
5 AGACCGGTGACGACGTCAACCGCGGAGTAGTACGCCTCCGCGCGGACGAC  
GTCGTCGAGGTGAGATTAGCGTCGGCTCCGACGACCAGGCGCATCGCGT  
15 CGGAGGTGAGGACGACGCCACCGTCGAGCGCGCGGCCCTCCGGCTCGTAG  
ACGTA CTGGAGGTGGTCTCTCTCGTCGTCACCGTCGAGGTGACGTG  
CCGCCGACGGCCTCGGGTCCACTCGATCGCGCGTCGTCGCGCGAGGGCTT  
20 CCTCGTACTGGGCCCACCACGCGCGCAGCTGCTTCGGCGTGCCGTAGCCC  
10 TCGGCCATGTGGGGTTCGAGCCCGCGACCTCGATGTCCACAGTCGGTA  
GAGGATCTGGAACGGCGTCATGGA CTTCGGGCCCCGGCCGTCTTGGAGT  
CCAGACGGCGGTCTCCATCGCAGCTGCGCCGGCGGCTTCGAGGTCTTGG  
25 TCGACGGAGTCGGGCGGCTCTCGCTTCCCGTCCTGGTTCTTGGTGAGGTA  
CTCGATCAGCGCAGCTCGTCAGCTGACCGGACGATCGAGACCATCACGC  
15 CGTGGCCCTTGCCCTTGCACTTGACGCCGGGGGTGTCGAGTCGGTCGAG  
GGCTCGAACTTGGGGTCAGCCCGCTTGAGGGCGCCCGCCACATCTCCCG  
30 GAGCCAGTCCTCCAGTCCCCAGGTCCGTCTCGGAGGGCTCAAAGTGTC  
CGACGACGTACCCCTTGGCCGGGGTGCCAGAGAGCTCGCCGCCGAGGAAG  
20 ACCAGCAGGTTGAGGTGGGGTGGTAACCGTTCTTCTTGGACCGGGTGAC  
35 CTCAGCCGCGCGGACCATGCCGATGTAGCCGATCCGGTGGCGGATGCCGT  
CCTCAGCGGGACGGACGTACTGCGTTCCGTCTTCCGGGTGCGGCGGGCC  
TCAGGGCGGCCGTAGAAGGCCGGGGCCGTGAGCATCCGCTGGTAGGCACC  
40 GGGCGCGCGGCGGGGCTTGCCCGACCGGTGAGGACCGGGGCGCCCTTGT  
25 CGTCCAGGAGAGGCCCGCCCCAGAGCGGGCGACCGGCTGTCGAGGTG  
GTGGTCTGGTTATGCCGGGCGGTGAGGACGACAACGGCGAGCGTGCCGCC  
GGCGGCGAGGTGCCGAGAGCACCGGTCTTGATCTCTCGGTCCGGCCAC  
45 GGCGGATCGCGGAGGAGCACTCCGGGCATAACAGATCCGCCCGCAGCGG  
ACCAGGCCGATCGTGACGACGTACCCACGGCTCGACTTCGCGTAGATCAC  
30 GCCGGTGTCCGGTTCGAGGACCCGTCGCCCGCACCCGCCGAGGCGTCGA

5

10

15

20

25

15

30

35

40

45

50

55

TCCCGGAGACCCGGTTGAGCACCTTGCGGCCCTGGTAGCGCGTACGGCA  
GCGGTGCGCCGCGCGCCTTGTGGTCTGTTCCGAAAGGGCTGCCGCCCTCTC  
GGACTCTCCCGTTCTTCCACGACTGCCACTTCCGCAAAGTCGCTGGTCA  
GTGGGGGGTGGGAAACTCTGTCAACCCTTTACCTAGGCGTCCCTTTTTTG  
CCAGGGGCGGTCTCACGGGCGGCCTCGGCGGCTCGGTGCGGGCCTTCCG  
GGCCCCGCGGGCCCGCTTCTTGACGCCTCGGAGCAGAACTCCTTGGCTC  
GCTTGCCGGGGGTGATCGTGAGGGCGGCGCCGACGCGCAACGGGGGCCG  
GCGGGACGCGGGCGGGCGACTGAGTCGGCGCGCCGATCAAAGAGGGGGT  
TGCGGACGCCAAAGCGTCCCTTACGCTGGACACAGACGAGTACCTTGGTT  
GGTAGCCGGGTGGACGTGAGAAGCGTCAGGGATTAGGACCCCTGGCCGT  
TTCGCTTTTTCTGGAGTTGTTTGGGTAGATCCTGCCGCATCGCCGCCTC  
ACGCGTGGCTCGCCGCGCGGATGCCTCAGAGGCCCCACCGGTCGTCAGGA  
CGCAGACGTGCGCGTGCTCCTGGTGGTGAGTCACCAGCTCGACCACACGG  
GCGCGGCCGGCGACGTCTC

SEQ ID No 77 (>Contig52)

CGGGATCTGGCCTTCATTAACCAACGACGGGGCAAACATAATAGGCTGGG  
CATTGCGCTTCAGCTCACCACAGCCCGTTTTCTGGGAACATTTCTGACGG  
ATTTAACTCAGGTTCTGCCTGGTGTTC AACATTTTGTGCGGTACAGCTT  
AATATCCACCGTCCAGAAGTTCTCTCCCGCTATGCTGAACGGGACACTAC  
CCTTAGAGAACATACTGCATTAATTAAGGAATATTACGGCTATCATGAAT  
TTGGTGATTTTCCATGGTCTTTCCGCTGAAGCGTCTGCTATATACCCGG  
GCGTGGCTCAGTAATGAGCGACCGGTCTGATGTTTGATTTTGCCACTGC  
ATGGTTGCTTCAAAATAAGGTATTACTGCCCCGAGCAACCACACTAGTAC  
GTCTCATCAGTGAAATTCGTGAAAGGGCAAATCAGCGGCTGTGGAAAAAG  
CTGGCCGCACTGCCGAACAAATGGCAGGCAGCTCAAGTGATGGAGCTTCT  
GGTCATTCCGGAAGGTCAGCGTGATCAGCACTGGAACAGTXXXXXXXXXX  
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXGCTGGAACGATAT  
ATCCGATTACGAAGTCTTGAGTTTCCCGACTGAACTTTCCGGTCTGCC  
TGCCATTCAACTGCGTAATCTGGTCTGTTATGCTGCCATGGCGTCGGTAA

5

150

10

5

15

20

10

25

15

30

35

20

40

25

AATATATCGCTCGAATGCCACAGCAGAGAAAGCTTGCTGTACTTACTGCA  
TTCGTTAAAGCACAGGAAATAACGGCATTAGACGATGCCGTTGATGTGCT  
TGATATGCTAATTCTGGACATTATCCGCGAAGCAAAGAAAACCGGGCAAA  
AAAAAAGACTCAGGACACTGAAAGATCTTGATCAGGCCGCATTGTTACTG  
GCGCGGGCATGTGCATTGTTGCTGGATGATAATACAGATGTCCAGATCT  
CAGGCAGGTTATCTTCAAGTGCCTACCCAAAAACAGACTGGCAGAATCTG  
TAAGCAAGGTTAATGAACTTGCTCGTCCACAGAACAXXXXXXXXXXXXXX  
XXXXXXXXXXXXXXXXXXXXXXXXXAAACGTTTTCTCCGGCGGTGTT  
GCGGGACCTGCATTTCCGTGCGGCACCGGCAGGTGAACATGTAAGGCTG  
CGATTCATTATCTGGCAGAACTGAATGGTTCGAAAAAGCGCATCCTTGAT  
GATGCGCCTGAACATATTATCACCGGTCCCTGAAACGCCTCGTATACGA  
TGCGGAGGGACGGATACAGCGTGCAGGTTATTCACTATGTTTGCTGGAAC  
GCCTTCAGGATGCACTGCGCCGCCGGGACATCTGGCTTGAAAACAGTGAT  
CGCTGGGGAGATCCTCGCGAGAAGTTGTTGCAAGGTGAAGAGTGGCAGAC  
TCAGCGTATTCTGTCTGTCTGGGCACTGGGACATCCTGTGATGGACGTA  
AAGGTGTGCAACAACTGGCTATTTCAGCTGGATGAGACCTGGAAAGCCGTG  
GCATCAGATTTGAAAAGAATGCGGAAGTTCATATCTGTAATGAAGGTAA  
ATATCCATCCCTGACTATCAGTTGTCTGGAGAAACAGGAAGAGCCACCAT  
CATTGCTTCGTCTAAATAATCGGATCAAACAGCTACTCCACCGGTAGAT  
TTAACGGAAGTGTACTTGAGATAGATGCCCAGACAGGATTTACACATGA  
GTTTGCGCATGTGAGAGAATCTGGTGCTCGAGCGCAAGATTTGCACATCA  
GTTTATGTGCGGTATGAATGGCTAAGCCCTGTAATATGGGCCTGAACCCG  
TTGATAAAGCACAAATATACCAGCATTGACCCGCCATCGGCTCAGTTGGGT  
GAAACAGAATTACCTTCGTGCAGAAACGCTGGT

SEQ ID No 78 (>Contig53)

45

30

ATTCCACGCGCTCACGGTCAGCTTCGACCCGCGCGAGCGCCCGGCGCCG  
CCTCGCAGAAGCGCGCGGTACGCTGTCCGAGCTCGGCGCGGACGCGCAG  
GCGCCGAGTGGCCGTTCTCGTCCGCGACGAGGCGGCGACCCGCGCGCT  
CGCCGAGGACCTCGGGTTCCGCTACGCCTACGATCCGACCACCGATCAGT

50

55

5

10

15

20

25

30

35

40

45

50

55

ACGCCCACCCGGCGGCCGTCTTCGTCTGACGCCGGACGGGCGGATCTCC  
CGGTACCTGTACGGGACGGAGTTCCCGGCGCGGATCTCCGGCTCGCGCT  
CCTGGAGGCGAGCCGCGGCGGTATCGGCACGATCGTCGATCGGGTGATCA  
TGACCTGCTATCGCTTCGACCCGGCGAGCCGGAGATACGCTCCGTTCCCTA  
5 CTCGGCTTCCTCCGGCTCGGGGCGGCGCCATCCTGATCACGGTCGGCGG  
GCTGCTCGCCGTCTGTGGCGGCGGAGCGCCGGCGGCCAGGTGCTCGCA  
CGAGCGCCGCGTCGGTCGTGACCCGTGGCCGACCGCCAGGGAGGTCA  
CCATGATCAACGAGCTCCTGCGCAAGCTTCTTTTCTGTCCGGCCAGTGG  
TCGACGATCGTGTTGACATTTACAAGCTGCTTTACTTCGTGATCTCGGT  
10 GACGATGGCCGGCGGACGCTCGTCGCCCTGTTCCGGCCTACCTGATGA  
TCCGGTACCGCAGGCGCCAGCGGGATGTTGAAGGCCCGTTCCCGGAGCG  
ACCGCGAGGCCTCCGCTCCTCCTCGAGGTCGGCATGGTGCTGGGCCTCAT  
CGTCCTGTTCTCTGTCGTGGTGGTCAATTGGAATGCGGCAGTATGCAGAGC  
25 TCCGCGTCGCCCCCGGACCCGGTCGTGGTGTACGTGACCGGGAAGCAG  
TGGATGTGGAAGTTCGCCTACCCGAGGGCCCGAGCTCGGTGGCGACGCT  
CTATGTGCCGGCGCGTCGGCCGGTGAAGCTCGTCATGACGTCCCGGGACG  
TGATCCACAGCTTCTTCGTCCCGATTTTCGCATCAAGTACGATGTCGTC  
30 CCCGGCCGCTACACCACGCTGTGGTTCGAGGCGACCGCGCCGGGCGCCTA  
TCAGATCCTGTGCACCGAGTACTGCGGGACGAACCACTCCACCATGCGCG  
GCGAGGTGATCGCGCTCGAGCCCTCCGATTTCCGCGGGTGGCTCTCCGAC  
20 CGCGGGCGGGGCGCCGGTATCGCCGGACAGGAGTACACGCCGCCGTGAC  
GCCGGGCGAGGGGATCCCGCGGAGCCGCTCAGCCTCGTCCGGCTGGGCG  
AGAACATCGCGGCCGAGGAGGGCTGCCTGCGCTGCCACACGCCGGACGGG  
40 ACACCGCACATCGGGCCGACCTGGGCCGGCCTCTACATGTCGGTCGTCCC  
GCTGGAGAGCGGCGGCGCCGGTCGCCGACGACGCGTACATCACCGAGT  
25 CGATGATGGATCCGCTCGCCCCGATCCACCGCGCTACCAGCGGTCATG  
CCCTCGTTCCTCGGCCGGCTCCAGCCGGCGCAGGTGCGCCCATCGTCGA  
GTACATCCGGTCGTTGAGGGGCGTCGCGCCGAGCCGGGCGCGCGGACGC  
45 CGCTGCCCCAGGGCCCGCCCTTCTGCGCTCCGGCCCCGAGCGCCCCGCC  
30 CCGCTCAGCGGGGCGCGCCGGTCGGCCCGATCGAGGGCGGCAAGCCCCG



5

152

10

GGAGGAGCTCCGATGAGCACGGAAACGTACGAATCTCTGCCCCGACGCGCC  
GGCCGAGAGGCCCGAGCCCCGACTACCTCCATGTTTACCGCGGGGTGACG  
GAGTGGCTCACGACCACGGATCACAAGCGGATAGGTCTCATGTTCTACGC  
CGTCATCGTCGGGAAAGCTTCTCTCTCGGAGGCATATTCGCCCTCATCAT  
5 GCGGACCGAGCTCCTCACGCCCCGAGCGGACCATCATCGACGCGGCGACCT  
15 ACAACCGGATGTTTACGCTGCACGCGGTGATCATGGTCTGGCTGTTTCATG  
ATCCCGTCGATCCCCAACGCGTTCGGCAACTTCGTCTCGCGATCATGCT  
CGGCGCCAAGGACCTCGCGTTCCTCCCGGATCAACCTCGCGAGCTTCTACA  
TCTACCTCCTCGGGGCGCGATCGCGATGGGCGGCATGATCGCGGCGGC  
20 10 ACGGACACCGGCTGGACGTTCTACCCGACGTACAGCCTGAAGACGCCGAT  
GACGCTGTTCCCGGTCGTCTTCGGCATCTTCATCGTCGGCGTCTCGTCCA  
TCATGACGCGGGTCAACTTCATCGTGACCACGCACACGATGCGCGCCGAG  
25 GGGCTCACGTGGAGCCGCTGCCGCTCTTCGTCTGGAGACCTACGCGAC  
GAGCATCATCCTGCTCTTCGCGACCCCGGTCTCGGGCTCTCGATCCTGC  
15 TCATCGGCATCGACCACGTGACCGCGCTCGGGATGTTTCGATCCCCGGTTC  
GGCGGCGATCCGGTCTCTTCCAGCACCTCTTCTGGTTCTACTCCCACCC  
30 CGCCGTCTACATCATGATCCTGCCGCGGTTCGGCGTGGTGAGCGAGGTCG  
TCTGACGTTTCGCGCACAAGCGCCCCGCGTCTACTGGGCGATCGCCATC  
TCGTGCTCGGGATCGCGTTCGTGGGTTCGTGACGTGGGGCCACCACAT  
20 GTTCGTGGCGGGGATGAGCGAGTACGCCGCGGACGTCTTCGGCGTGCTCT  
35 CGATGTTTCGTGGCCATCTTCTCGGCCATCAAGGTCTACACGTGGGTGCGG  
ACGCTGTACAGGGGCTCGATCCACTTCAACACGCCGCTGCTCTACTTCAT  
CGCCTTCCTCTTCTGTTCTGTTCTTCGGGGGATGACGGGCGTGGCCGTCG  
40 CCACGCAGTCGCTGGACGTGCACTGGCACGACACATACTTCGTTGTGGCG  
25 CACTTCCACTTCATCATGGTGGGCGGGACGCTCACCATGTTCTTCGCGGC  
GGCGCACTACTGGTTTCCGAAGATGTTTCGGGCGCCTCTACTCGGAGCGCG  
TCGGGCTCCTCTCGGCCGCGTCGGTGTCTCCTCGGCTTCTTGTGACCTTC  
45 TTCCCGCAGTTCTCTCCTCGGGAACATGGGGATGCCCCGCGCTATTACAG  
CTACCCGCGCGCTACCACTGGCTCCACGTGCTCTCGACGGCGGCGCCT  
30 ACCTGCTCGCCGCGGCGCTCGTGAATCTCGCTCCTGAACCTCGTCATCGCG

50

55

5

10

15

20

25

30

35

40

45

50

55

CTCAAGTGGGGCCGGAAGGCCGGGAGGAACCCCTGGGGCGGGCGCACGCT  
CGAGTGGATGACCGGCGAGCCCTTGCCGCCCAAGCACAACTTC<sub>5</sub>CGGTTCG  
CGCCGCTCGTCCGCCGCGGCCCGTACGAGTTCCAGCTCTCCGAGGAGGAC  
GCCCCGTGCGACAACCACGCCCGCTGCGTGAGCAGTTCGAAGATCTCGAGA  
AGCAGACGCACGCGGCCCGCTCGGGATGTGGTTGTTCTCTCGGGAGCGAG  
GTGCTCCTCTTACC<sub>10</sub>GGGCTCTTCGCGCTGTACGCGGCGTACCGCGAGCT  
CTACCCGCGCGATTTCCGCCGAGGCCATCGCGCACAAACGTCGCGATCG  
GCACCACCATGACGCTCATCCTGATCGGCAGCAGCTTCACCGTCGCCATG  
GCGGTGCACGCCGTC<sub>15</sub>CGCGCCTCCCACCCGCGCGCGCGCGCTGTTCTT  
CGCGGTGAGCGTGGCGATCGGGATCGTGTTCTCTCGTGCTGAAAGGGATCG  
AGTACGCGCAGCACTTCCGCGAGGSCATCTTCCCGCGCGCGCCTACCGC  
TTCGCGGAGCTCCCGACGTTCCGGCGCGCAGATGGCGTTCACGCTGTACTT  
CGCCATGACGGCGCTCCACGCCCTGCACGTCGTGGGCGGGGCGGCCCTCC  
TCACGGGGGTGCGGTGGGGGTGCTGGAAGGGCCGGTACTGGGCATACGAC  
CAGACGCCCGGTGGAGCTCAGCGGCCCTTACTGGCACCTCGTGGACATCAT  
GTGGATCTTCACTGGCCGCTCCTCTACCTGACGCGCAAGTGACGGCACA  
GCCCCGAGACGACCATGCCGCAAGAGCACGTCGCGGAAAGCACGCCCTGG  
ACCCGTTACCTCATGGACGCTGATGSCCCTCATTGCCCTCACGCTCCTGT  
CGTTAGCGCTCTCGTTCTTGCGCACGGGGGCTTGGGAAATACCGATCGCG  
CTGCTCCTCCCCGTGGTGAAGAGCGTGCTCGGGCTTGCTTTCTTCTTGCT  
CCGTGAGGTGCATTAAGGTCATCAAACGCCTTTTCAAACATTGGCTGC  
TGGTGTACTTACTTGGACGTTTCATACTGTCTCTTTATGGCCCCCAACG  
GTCATTACGTCGACCTACACAATTTCTCTTCCCGCACCCCTCCTACCTGTA  
TCTCTAAGCACTGCCTTGCGTCCTGCTCTATTACATTCTACTCCGGCTGT  
CCATGTGTGGGATTATATGCGCGAGGTACCTATTCCGCCGTGGAGTCTCC  
ATTTACCTCTTGGACCTTGCCCGTCTGAATCTCGATCTCCTCATGCGTT  
GGTCCACCATGTATTACCTCCTAGAACTTTATACTCCATATCTCTATATA  
TCTAGTTGTGCGTGTAATTGTGTGCATATATTATCGCCACTGTGTATGAA  
TACCGTGCCGACGTGCTATATACGAAANTACTCCTCGGTGATATCTCCA  
CCTCATATATACCTCCGAGTGTAGTATACGCACGAGTGTATATACTCTTC

5

154

10

CTCTGGTCACGCGACTTCGTGCTGATATGATACCATCGTTCCATGTTACG  
CGAAGTTACTCATAAGATCTCCTCACACATCAACGAGTGTACTCCTATGT  
GTTTCATACAAACTCGATACCCTTCAGAGTAGTGTGCCTATGTGGTA  
TGCATAATGTTAGTATACTTT

5

15

SEQ ID No 79 (&gt;Contig54)

20

10

TGGGAAAGAGGGCCACAGGGGATGTAGCAGGACGCTTAATAGTAAATGAC  
GAGGGTGTGCCGACGAGACCCGTAGGAAACAACGGGCACAGACGAGAGCA  
ATAAAGGGGGTTTGAAGGTACCCCGGATAGAGTAGAGAAGGCTAGCGGAC

25

15

GAGTAAGACGCGGAGGAAATAAGTCGGCGTCGTAGAAGTTCTGTGGAGAA  
GGTACGACTCTTAAAGACCTAGGCGGGAGACAGTTTCCACCCGAGGCAGA  
GCAAGACCACAAGATTCAGAGGGAGTAAGGAGTTCGGAATTGGAGAGGTT  
GAGGGGCGTGTGAGCCGTCAAGTGGGGCGGTACGCAAAGAAAGAAGCGT  
CCATGTCAGAGGCCCAGCGCCGTTGCGGCCCTACATGGGTGATCACGGT

30

20

GGGCCGTAGGCGGACCGGAGATGAGCGGGCTCCGCCACCGGACGGCGA  
GGGGCACGGCGCCTTCGTCCGGGGCGCTCGCGCGCGATCTCGCGCGCGCG  
GGGGGCGCCGGCGTCGCGCTCGTCGGCGACGGCCAGCCGCCCATCGTCCA  
CGCCCTCGGGACGTCATCAACGCCGCGCTCCGCAGCCGGGCGGCCTGGA  
TGGTCGATCCTGTGCTGATCGACGCGGGCCCTCCACGCAGAGCTTCTCC

35

25

GAGCTCGTCGGCGAGCTCGGGCGCGCGCGGTGACACCTTGATCCTCCT  
CGACGTGAACCCCGTGTACGCCGCGCCGGCGACGTGATTTCGCGGGCC  
TCCTCGCGCGCGTGCCACGAGCTTGAAGGCCGGGCTCTACGACGACGAG  
ACCGCCCGCGCTTGACGCTGGTTCTGTGCCGACCCGGCATTACCTCGAGTC  
GTGGGGGACGCGCGGGCGTACGACGGGACGGTCTCGTTCTGTGCAACCCC

40

30

TCGTCCGGCCGCTGTTGACGCGCCGGGCGGTGCCCAGAGCTGCTCGCCGTC  
TTCGCGGGGACGAGCGCCCGGATCCCCGGCTGCTGCTGCGCGAGCACTG  
GCGCGGCGAGCGCGAGGGGCGGATTTCGAGGCCTTCTGGGGCGAGGCAT  
TGAAGCGCGGCTTCCTCCCTGACAGCGCCCGGCGAGGCAGACACCGGAG  
CTCGCGCCGGCGATCTCGCTCAGGAGCTCGCGCGGCTCGCCGCCGCGCC  
GCGGCCGGCCGGCGCGCGCTCGACGTGGCGTTCTCAGGTGCGCGTCGC

50

55

5

155

10

5

15

20

10

25

15

30

35

20

40

25

45

30

50

55

TCCACGACGGCAGGTTCCGCAACAACCCCTGGCTGCAAGAGCTCCCGCGG  
CCGATCACCAGGCTCACCTGGGGCAACGCCGCCATGATGAGCGGGCGAC  
CGCGGCGCGGCTCGGCCTCGAGCGCGGCGATGTCGTCGAGCTCGCGCTGC  
GCGGCCGACGATCGAGATCCCGGCCGTCGTCGTCGCGGGCACGCCGAC  
GACGTGATCAGCGTCGACCTCGGCTATGGGCGCGACGCCGGCGAGGAGGT  
CGCGCGCGGGGTGGGCGTGTGCGCGTATCGGATCCGCCCGTCCGACGCGC  
GGTGGTTCGCGGGGGGCTCTCCGTGAGGAAGACCGGCGCCACGGCCGCG  
CTCGCGCAGGCCAGCTCGAGCTCTCCAGCACGACCGTCCCATCGCGCT  
CAGGAGGACGCTGCCGAGTACCGTGAACAGCCCGGTTTCGCGGAGGAGC  
ACAAGGGGCCGCTCCGCTCGATCCTGCCGAGGTCCAGCACACCGGCGCG  
CAATGGGCGATGTCCATCGACATGTGATCTGCACCGGGTGCTCCTCGTG  
CGTCGTGGCCTGTGAGGCCGAGAACACGTCCTCGTCGTGCGCAAGGAGG  
AGGTGATGCACGGCCGCGAGATGCAGTGGTTGCGGATCGATCAGTACTTC  
GAGGGGGGAGGCGACGAGGTGAGCGTCGTCAACCAGCCGATGCTCTGCCA  
GCACTGCGAGAAGGCGCCGTGCGAGTACGTCTGTCCGGTGAACGCGACGG  
TCCACAGCCCCGACGGCCTCAACGAGATGATCTACAACCGATGCATCGGG  
ACGCGCTTTTGTCTCAACAACCTGCCCGTACAAGATCCGGCGGTTCAATTT  
CTTCGACTACAATGCCCACGTCCCGTACAACGCCGGCCTCCGCAAGCTCC  
AGCGCAACCCGACGTGACCGTCCGCGCCCGCGGCGTCATGGAGAAATGC  
ACGTACTGCGTGACGCGATCCGAGAGGCGGACATCCGCGCGCAGATCGA  
GCGGCGGCCGCTCCGGCCGGGCGAGGTGGTCACCGCCTGCCAGCAGGCCT  
GTCCGACCGGCGCGATCCAGTTCGGGTGCGTGGATCACGCGGATACCAAG  
ATGGTCGCGTGCGCAGGGAGCCGCGCGGTACGCCGTGCTCCACGACCT  
CGGCACCCGGCCCGGACGGAGTACCTCGCCAAGATCGAGAACCCGAACC  
CCGAGATTGAATGAGCCATGGCGGGCCCGCTCATCCTGGACGCACCGACC  
GACGATCAGCTGTGGAAGCAGCTCCTCGAGCCGGTATGGAAGCCCGCTC  
CCGGCTCGGCTGGATGCTCGCGTTCGGGCTCGCGCTCGGCGGCACGGGCC  
TGCTCTTCCTCGGATCACCTACACCGTCCTACCGGGATCGGCGTGTGG  
GGCAACAACATCCCGGTGCGCTGGGCCTTCGCGATCACCAACTTCGTCTG  
GTGGATCGGGATCGGCCACGCCGGGACGTTTCATCTCCGCGATCCTCCTCC

5

10

15

20

25

30

35

40

45

50

55

TGCTCGAGCAGAAGTGGCGGACGAGCATCAACCGCTTCGCCGAGGCGATG  
ACGCTCTTCGCGGTCTGTCCAGGCCGCTCTTCCGGTCTCCACCTCGG  
CCGCCCCCTGGTTCGCCTACTGGATCTTCCCGTACCCCGCGACGATGCAGG  
TGTGGCCGCGAGTTCCGGAGCGCGCTGCCGTGGGACGCCGCCGCGATCGCG  
5 ACCTACTTCACGGTGTCTGCTCCTGTTCTGGTACATGGGCCTCGTCCCGGA  
TCTGGCGGCGCTGCGCGATCACGCCCGGGCCGCGTCCGGCGGGTGATCT  
ACGGGCTCATGTCTGCTCGGCTGGCACGGCGCCGCGACCACTTCCGGCAT  
TACCGGTGTGTACGGGCTGCTCCCGGGGCTCGCGACGCCCTCGTCTGT  
CTCGGTGCACTCGATCGTGAGCAGCGATTTCCGGATCGCCCTGGTCCCCG  
10 GCTGGCACTCGACGCTCTTTCGCGGCTTCTTCTGTCGCGGGCGCGATCTTC  
TCCGGGTTCCGCGATGGTGCTCACGCTGCTCATCCCGGTGCGGCGGATCTA  
CGGGCTCCATAACGTCTGACCGCGCGCCACCTCGACGATCTCGGAAGA  
TGACGCTCGTGACCGGCTGGATCGTTCATCCTCTCGTACATCATCGAGAAC  
25 TTCTCTCGCTGGTACAGCGGCTCGGCGTACGAAATGCATCAGTTTTTCA  
GACACGCTGCGCGGCCCCGAACAACGCGCCTACTGGGCCAGCACGTCT  
15 GCAACGTGCTCGTCATCCAGCTCCTCTGGAGCGAGCGGATCCGGACGAGC  
CCCGTTCGCGCTCTGGCTCATCTCCATCCTCGTCAACGTCCGGATGTGGAG  
CGAGCGGTTACGCTCATCGTGATGTCTGCTCGAGGAAGAGTTCTCCCGT  
CCAAGTGGCACGGCTACAGCCCGACGTGGGTGGACTGGAGCCTCTTCATC  
20 GGGTCAGGCGGCTTCTTCATGCTCCTGTTCTGAGCTTTTGGCGCTCTT  
TCCGTTTCATCCCCGTGCGGAGGTCAGGAGCTCAACCATGAAGAGCTGG  
AGAAGGCTCGGGGCAAGGGGGGGCTGATGGAGACCGGAACGCTCGGCG  
AGTTCGACGACCCGAGGCGATGCTCCATGCGATCCGAGAGCTCAGGCGG  
40 CGCGGCTACCGCCGGGTGGAAGCGTTTACGCCCCATCCGGTGAAGGGGCT  
25 CGACGAGGCGCTCGACCTCCCGCGCTCGAACCTCAACCGGATGGTGCTGC  
CCTTCGCGATCTGGGGGTCTGCGGCGCTACTTCGTCCAGTGGTTCTGC  
AACGCTTTCCACTATCCGCTGAACGTGGGCGGGCGCCGCTGAACCTCGGC  
45 GCCGGCGTTTCATCCCGATCACGTTTCGAGATGGGGGTGCTCTCCACCTCGA  
TCTTCGGCGTGCTCATCGGCTTTTACCTGACGAGGCTGCCGAGGCTCTAC  
30 CTCCCGCTCTTCGACGCCCCGGGCTTCGAGCGCGTCACGCTGGATCGGGT

5

157

10

5

15

20

10

25

15

30

35

40

25

45

30

50

55

CCTGGTCGGGCTCGACGACACGGAACCTTCCTTCTCGAGCGCCAGGCGG  
AGCGCGATCTCCTCGCGCTCGGCGCCAGGCGCGTCGTCTGGCGAGGAGG  
CGCGAGGAGCCATGAGGGCCGGCGCCCGGCTCGCCCCCTCGGGCGCGCG  
CTCGCGCCGTTCGCCCTCGTCCTGCTCGCCGGGTGCCGCGAGAAGGTGCT  
GCCCCGAGCCGGAATTTCGAGCGGATGATCCGCCAGGAGAAATACGGGCTCT  
GGGAGCCGTGCGAGCACTTCGACGACGGCCGCGCGATGCAGCACCCGCCC  
GAGGGGACCGTCGCGCGGGGCGCGTCACCGGGCCGCCCGGCTATCTCCA  
GGGCGTCTCTCGACGGGGCGTACGTACGGAGGTGCCGCTCTCGCTCACGG  
TCGAGCTCGTGACGCGCGGCCGCGCAGCGCTTCGAGACCTTCTGCGCGCCG  
TGCCACGGGATCCTCGGCGACGGCAGCTCGCGCGTGGCGACGAACATGAC  
GCTGCGCCCCGCCCGCTCGCTCGTCGACCCGAGGCGCGGAGCTTCCCGC  
CGGGCAGGATCTACCAGGTCAATCATCGAGGGCTACGGCTGATGCCGCGC  
TACTCGGACGATCTGCCCGACATCGAAGAGCGCTGGGCCGTCTGCGCCTA  
CGTGAAGGCGCTTCAGCTGAGCCGCGGAGTGGCCGCGGGCGCCCTCCCGC  
CCGCGCTCCGCGGCCGGGCAGAGCAGGAGCTGCGATGAACAGGGATGCCA  
TCTAGTACAAGGGCGGCGCGACGATCGCGGCCTCGCTCGCGATCGCGGCG  
CTCGSCGCGGTGCGCCGCGATCGTCGGCGGCTTCGTGATCTCCGCCGTT  
CTTCTTCTCGTACCTCGCCGCGTGGTCTGTCGCGTATTCTGTCCGTGG  
GCGCGCTCGTCACGCTCCTCACCTGCAACGCCATGCGCGCGGGCTGGCCC  
ACGGCGGTGCGCCGCCTCCTCGAGACGATGGTGGCGCCGCTGCCCTGCT  
CGCGCGCTCTCCGCGCCGATCCTGGTGGCCCTGGACACGCTGTACCCGT  
GGATGCACCCGAGCGGATCGCCGGCGAGCACGCGCGGCGCATCCTCGAG  
CACAGAACGCCCTACTTCAATCCAGGCTTCTTCGTCTGCGCTCGGCGAT  
CTACTTCGCGATCTGGATCGCCGTGCGCCCTCGTGCTCCGCCGGCGATCGT  
TCGCGCAGGACCGTGAGCCGAGGGCCGACGTCAAGGACGCGATGTATGGC  
CTGAGCGGCGCCATGCTGCCGGTCTGGCGATCACGATCGTCTTCTCGTC  
GTTGCACTGGCTCATGTCCCTCGACGCGACCTGGTACTCGACGATGTTCC  
CGGTCTACGTGTTTCGCGAGCGCCTTCGTGACCGCCGTCGGCGCGCTCACG  
GTCCTCTCGTATGCCGCGCAGACGTCCGGTTACCTCGCGAGGCTCAACGA  
CTCGCACTATTACGCGCTCGGGCGGCTGCTCCTCGCGTTACGATATTCT

5

10

15

20

25

30

35

40

45

50

55

GGGCCTATGCGGCCTATTTCCAGTTCATGTTGATCTGGATCGCGAACAAG  
CCCCACGAGGTCGCCTTCTTCCTCGACCGCTGGGAAGGGCCCTGGCGGCC  
GACCTCCGTGCTCGTCGTCTCACGCGGTTCTGTCGTCCCGTTCCTGATCC  
TGATGTCGTACGCGATCAAGCGGCGCCCGCGCCAGCTCTCGTGGATGGCG  
CTCTGGGTGCTCGCCTCCGGCTACATCGACTTTCAGTGGCTCGTGGTGCC  
GGCGACAGGGCGCCACGGGTTGCCTATCACTGGCTCGACCTCGCGACCC  
TGTGCGTCGTGGGCGGCCTCTCGACCGGTTCCGCGGTGGCGGCTGCGA  
GGGCGGCCGGTGGTCCCGGTCCACGACCCGCGGCTCGAAGAGGCCTTTGC  
GTACCGGAGCATATGATGTTCCGTTTCCGTACAGCGAGGTTCCGCCAGGA  
GGAGGACACGCTCCCTGGGGGCGCGTGATCTCGCGTTCGCGGTCGTGC  
TCGCGATCGGCGGCGCGCTGACGCTCTGGGCTGGCTCGCGATGCGGGCC  
CGCGAGGCGGATCTGCGGCCCTCCCTCGCGTTCCTCGAGAAGGATCTCGG  
GCCGCGGCGCGAGGTCAGCATGGTCCAGCAGTCGCTGTTGACGAGGCGC  
GCCTGGGCCAGCAGCTCGTCGAGGCGCAGCGCGGAGCTCCGCCGCTTC  
GGCGTCGTGATCGGGAGAGGGGCATCGTGAGCATCCCGATCGACGACGC  
GATCGAGCTCATGGTGGCGGAGGGCGCGGATGAGCCGGGCCGTGCGCGT  
GGCCCTCTGCTGGCAGCCGGCCTCGTGTGCGGCCCGGGCGCGCTCCG  
AGCCCGTATCGCTTTCGCCCCGCGTGGGCCCGTCCGCGGGCGAGGCCGC  
GCTCTGAAACGACGGCTCCGGCGCGGATGAGCGGCCCGAGGCGACCTCCT  
GCAACCCACCGCGCTGCGTACAGGGGTAAATCAACTGCATTCCAGGATAC  
GGACCGCGCAGCATAAACCCCTCACCGGACAGACTGAATAATGCCGAACCT  
TGAACTTCTATGCCCATGCGTGCGGGATCATCATGCCCAACTATTTAAAC  
TGGTCCCTCCGGCAAAAGGAACGGACCAGCACCAGAATAAACCCTGTTT  
GGCCAGCGCAAAAAATGAAAACTCTTCTCCTTGCGGCTAAATAAACGA  
CTCCGGGAACCGAATGGTATGCAATACACCATGGCAACGCGATTCTGTC  
TAACCTTAAACAACTTCCCAGCGACTCGTCCGTGCAACGCTTCTGACGC  
AAGACCATGACCCACGAGAACCGGGCGGCGACACTGCCAGTGAAAC  
TCGCCTAGGCCCGCCCTGCCTTCACGTATTCACCGTGGGGCGGTCCAAA  
ACTAAAAACAATACGTGACTCTTCTCAAATATCGTCGGATAAAGGCCAAC  
ACGCGTACCTCCCCCTAAAGGGAAGAAACCCCTACCAGGGTGGACCGTAT

5

10

CCACCCGTGATCCCTGAAACATCTCCAGCCGCGTACAAATTAGGCTTTGA  
CAAAACC

SEQ ID No 80 (>Contig55)

15

20

25

30

35

40

45

50

55

5 GGGGAAAGAAGTCGGGAGCAGCAGAAGAGCGCAGCGATTGAAAGCACGAC  
15 GCAAGGGCAAAGGACATNAGACGACAACAGAAAGGCGAAGACATGAGAAG  
AGGAGAAAGACTCGGAAATGGGCACAAAGCGAGAAAAACGTAACGCTAT  
CTGAGAAAGAACACCATAAAGTCAGACTGGGGTAAAGCCATACACGCAGC  
GAAAGCAGACAAGCAAAAGTCATTAACAGATGCAACAAAGAACAAAAGAG  
20 10 AAAGGAAGGACACATGAAGGAGAAAGCGGCTTCGAGAACTAAAGACGGG  
AGCCAACAATATACCTTTCACAGGGSCGAAGAAGGGCCCAAGGTTCAATC  
GAAGACGATTGAATCCAAGGAAGTCCAATCGGAAGAAAAAGCATATTGA  
TAAACCAGACAACGACAGCAGGCCTGAACGTAGGAGCGAGATCGTGAGAC  
25 ATCAGTAGGCAAAACAAGAGCGCTACACCCAGGGTCTGTC AACCTAGAAA  
15 GCGCGCTCCTCAAGCCGGTAGCGGCCGCGCGCACCAGCCCGATGCGGGC  
GCCTTCGCGCGCGAACC GCCCGCACCGTTCGCGCGCCCCACGCCCGCTGACG  
30 CCCCCGTGATCACCACGACCTCCGGCCGCCTGGAATCGCCCATGCCCGTC  
GCCTCCCGCTCGCCCGCGGCGTCAAGCAACGTGAATGCCACCTGAGCGT  
GTCACTTCCTCAAGCTCGACAGCACGTCCTTGATCCGCTCGGTCCGACCC  
20 GCCGTGAGCGACCAGGGGCTCGGCTGCCGCACGACGTGCAGGTGCGCTCG  
35 ACGCTCGAGCACCTCGAACC CGCTCGTTCCCTGCTCGGTCCGACGAACC  
GGAGCGCCACCACGGCCGTGCCACATGGAGGTTGAGAGGGTGATCTCG  
GGCAGCCACGCCGCGCAGCGCCGGTCCGACCAGCAGCAGGTGAGCGGCGC  
40 GAACGGGTACAGCCCAGCATCGCCTGGAGCATCGGAACACGCTCGAGC  
25 AGGACCAGGCCTGCGGCCAGTTGGCCTTCGGGTACAGCGGGGAACGGG  
TGCTCCGCGTCGCGCGGGTGGCCGCTCCAGCACTCGGGCAGCCGTTGGTG  
CTCGAACAGCGCCCGCCCTCGAACACGGCGCGGCACAGGAGCGCCACGT  
45 GCCCGTGCAGGCCGTACCGCGGAGCCCGAGCGCGATCGCGCCCTGATCG  
ACGGGCCAGACCGTCCCCCGGTGATAGCTGTACGGATCGAACGCGGGGTG  
30 ATCGGCCGAGAGCGTGCGGATACCCAGCCGGAGAACATGTCCGCCGCGA



5

160

10

15

20

25

30

35

40

45

50

55

ACATGCGGATGGCGGTGGGCTCGGCGAGGGCCTGGTCCACGATCCCCGCC  
GCGAGGCAGAGCCCCGGATCCGAGCCGATCGAGCGGATCTGGCGCTTGTC  
CGGGCCGAGGCCCCATCGCGAAGGTGCGCGCGTCGGGCATCCAGAAGGCGT  
CGTTGAACCGCCTCTGCAGCTCGAGCGCCTCGGCGAACAGCCGTCGCGCG  
5 TCGTCCTTGGCGCCGAACCAGAAGAGCAGCTCGGAGAGGCGCAGCTTCGA  
CAGGAACACGAAGCCCTGCATCTCGCACGTCCCGATCGGCGGCCGCACCT  
GAGAGCCGTCGGCGTGGACGATGGCGTCGTGGAGTCCCTCCAGCCCTGG  
TTCTGGATCGACGCGCTCGAGCGGGGCTCGTACTCGTAGAACCCGTCGCC  
GTGAGATCGCCCTCCTCGTCGATCCAGCGCATGGCCCTGAGCGCAGGCT  
10 CGATCAGGCGGCCGACGCGCTCGCGATCGCCGGTCCAGTGCCAGAGCTCC  
GAGGCGGCCACCGCGTAGAACATCGTCGAGGTGCGCGACGCGTACGTGCG  
CCCCAGCGGGTTGTAGTTGAGGTCCGAGAGCGCTCCGTCCCTGGCCTGAT  
GCAGCATCCGGTCGGGCTGCTCGTCGCGCCAGTCGTGACGACGCGCCCC  
TGCCAGCGCGGGAGCACGAGCGCTGTGCCTGCGAGGATGTGGTTCGTACG  
15 CGCGCGGCCTGCGTGCCCGCGGCGAGCGGGTCGCGGCCGAAGAGCCGA  
TGTAGATCGGCAGTCCGGCGGCCACCGTCCAGGAGCGCTCGTCCTGGTCG  
ATATCGTACATGCGGAGCGCGATGAGATCGCGCTTGCTCGCTCGAGCAC  
GGAGAAGACCGTATGCGAGAGCGTGTGCGCGCCCGGGACCGAGAAGGACG  
TGGCGCGGTGGTGAAGGTGCCGCGGCCACGTGCGCGCGTTTCGCGTG  
20 CCGAAGAACGAGCGGCAGCCGGCGAGCAGCGGGAGCGGTTCCGCCCGGAT  
CAGGGCGATCACGTGACGCGAGCCGCGCCAGGTGCCATGGGGCTCGAGAT  
CGATCGAGAACCAGATCTCGCGGCCAGCGCAGGACGGGGGCGAGCCCGCG  
CTCCGGGCCCCGTACGACGATCCCGGCGTCGAAGCGCGCGACCCGGATTCT  
TCCGGGGTGTTTCATAGCGGTGCTCGGCGCGGTAGTCGACGCGCAGCGCGC  
25 ATCCGTGCTCGGCGGGCTCGAGCGTCCACCGGGCGTCGCCGCGCTGGAGG  
CGCGGGCCGTCCGTCTCCTCCGCGTCGGCGAAGTCGGCGTCGATCTCGAG  
CGCCAGCGTGAAGCGCACGCGCTCCTGCGTGAAGCTCGCGACGTCGATGT  
CCTCGTGGAAGCCGTCCCCACGAAGCGCGACAGCCGAGCTCGACGGTC  
45 CGCGGCGCCGCGCCCGCCCTCGGGCGGCGGGGCGATGTAGTAGCCGAG  
CCAGCTGTCCGGCTCGACGGCGGAGAGCGCGACGGGCGCCGGCGGCCTGC

5

10

15

20

25

30

35

40

45

50

55

CGTCGATCAGGTGGCGGTAGAGGGAGAGGAGGCGGGTGTTCGCGACGAAC  
AGGCCGATGTGCGGCTCGGGGGCGATCGAGCCGTCCGGACGCATGCACAG  
GACCGTCCGGTCTGGCTGATGTACAGGGAGCCGGCGGGGGCCTCAGCG  
TGGCCAGCGAGCCGAAGGGTCTCTCGAGCGACATGGTCACCTCCAGGCGA  
5 GCGCGGGGTGAGGACTGCCAGGGGCGTGCCAACGACGTGGACGCGCTCAC  
GCCAAGCCGTGGGCGGGCGGCGGGCCCCCTGCGCCGAGGCTCAGCCGA  
15 GCGCGAGCGGTGTTTCGTGCGCGGAGCGGGCCTCGTCGTGCGCTCCAGG  
TAGGCCTTTCCGATGTGCGCCTCGTCCACGAACCCGACGATCGCGCCCTC  
GCGCTCGACGACGGGGACCTCGCGGACGCCGTGCGCCACCATCGCCTCGG  
20 TCGCCGTCCGAGATCGTCGGTGACCGTCACGGCCACCGCGGCTGCATC  
GCGTCGGCGGCCACGGTCATCCGCTCGAGGTGCTGCTCCACCGCGATGAT  
CCGGAGCGACTCGGCGGTGATCATGCCGACCATCTTGCGCGACGGTTCGA  
25 GCACCGGGAACACCTCCTGCCAGCTCGCGTCGGCCGCCCCCGGAGCATC  
TCGCGGGCCGGCGTCCCCGGCACGAACGTACAGAGCGCGCGCCCTCGAT  
15 CATGATCTGCCGCACGCGGATGGTCTTGAGCACGTCGAGCGTCGGCACCG  
GGTGC GCGGGAGACTCGCGCTGGGTGGGGAGCTGCGCGTGGTAGAGCGAG  
30 TGCTTCCGAGCGCGACGAAGGCGACGCCCTCGGCGAGCATCAGCGGAAC  
CAGGAGGTGCTAGCTGCCGGCGAGCTCGCAGACCATCAGAGGGAGCTCA  
CCGGCACGTGCGCGACGCCCGCTAGAAGGTGCCCATGCCACGAGCGCG  
20 AAGCGCCCCGGTTCGATCGCGCGATCGCCGAGCAGGAGCGCCGCGCGCG  
35 CCCGAACGCGCCGCCGAAGAGCCCGCCGATGACGAGCGACGGCGCGAAGT  
CGCCGGCGCACCCGCCGCTGCCGAGCGTGAGCGACGAGGCGACGATCTTG  
GCGCGCAGAGCAGGAGCAGGAGCTCCACGCCGCGCCAGCCCGGTGGAG  
40 CCACGTGGCGCCGGTGATCGCGACCTGGACGGCGCCGTACCCGCCGCCGA  
25 GCAGCCCGAGCCCTTGCCCGGGGCTCTCGATCCTGCGGCCGACGAACCAG  
AGGACGGGCACGCAGAAGAGGCCAGCGCGAGCCCGGAGCCCCGGGCG  
CGCCAGGGGGCGATGGGCAGGCGCGCCGCGATCGCCTTCACGCCGCCGA  
45 GGCACTTCAGGAAGCCGATCGCGACGAAGGCCAGGAGCAGCGCGAGCAGC  
GCATAGAGAGGGAGGTGCGACGGGACGAACGCATACTTCGGCGCGTGCGC  
30 GAAGAGCGTCGACTCGCCGTAGAACGAGATGAAGACCGAGTAGGAGACCA

5

162

10

CGCTGGCGAGCAGCGCCGGGATCAGCGCCTCGGCCTCGAAGTCGTGCGGG  
TAGAGCACCTCGACGGCGAGCAGGGCGGCGCCGAGCGGCGTGCGGAAGAT  
GGCGGACATCCCGGCCGCGACCCCCGCGAGCATCAGGATGCGGTGCTCGC  
GCCGGCCGACCGCGAGCCCGCGCCCCACGAGCGAGCCGAGCGCTCCTCCG  
5 ACCTGCATGGTCGGCCCCCTCGCGACCGCCGGCGCCGCCGAGCCGAGCGT  
15 GAGGATCGACGCGACCGCCTTGACCCACGCGACCCGCTTGCGCATCCGGC  
CGCCGTGGTGGTGAAGGCCTGGATCATCGCGTCGCCGCCGCCGCCCGCG  
GCCTCGGGGGCGAGGCGCCAGGTGAGGATGCCCCCGGCCAGCGCTCCGAG  
CGCCGGGATCAGCAGCAGCAGCCAGAGCCGACGCTCCGGTGCTCGGGGC  
20 CGTCGCCCGCTGAAGATGGCCTCGCCGTGGGCTCGAAGCCGCGCGTAGCCG  
10 GCGAGGCGGGCCGAGGAGCAGCTCCTCGACGAGCTCGAGGGCGCCGAAGAA  
GAGCACCGCGACGAGGCCCGGATGGCCCCGACGAGCACGGCATGCAGGA  
25 TCGTGCGCCCGACGAGCCTGAGATCGAGGGGCGCGACCTCCGAGAGGAGC  
GCGGAGAAGGGGCGCCGCCGTGCGACACGCCGGCGCGCTCTGCTGGGAA  
15 TGGTTCCGTACGGCGATGGT

30

SEQ ID No 81 (&gt;Contig56)

35

GGATCCGGCCGCGAGCGGCTGCGGTGCCGATGCGCTCGACGGTGACGGGC  
GGGGTGATCGCGGGTCCGGAGCTCGGTGCGAGCTACTGGGCGGACAACT  
20 TCGGCAGCCGGTGCGCTTCGCTGCGGGCGCAAGCGCTGCTGGAGGGTG  
GCCCCGCGCTGTTTCATCGAGATGAGCCCGCACCCGATCCTGGTGCCGCCC  
CTGGACGAGATCCAGACGGCGGCCGAGCAAGGGGGCGCTGCGGTGGGCTC  
GCTGCGGCGAGGGCAGGACGAGCGCGCGACGCTGCTGGAGGCGCTGGGGA  
40 CGCTGTGGGCGTCCGGCTATCCGGTGAGCTGGGCTCGGCTGTTCCCCGCG  
25 GGCGGCAGGCGGGTTCCGCTGCCGACCTATCCCTGGCAGCACGAGCGGTG  
CTGGATCGAGGTCGAGCCTGACGCCCCCGCCTCGCCGAGCCGACCCCA  
CCAAGGACTGGTTCTACCGAACGGAAGTGGCCCCGAGGTGCCCCGCGCCGCC  
45 CCGAAATCGGAGACAGCTCATGGGAGCTGGCTGCTGTTGGCCGACAGGGG  
TGGGGTCGGTGAGGCGGTGCGTGCAGCGCTGTCGACGCGCGGACTTTCCT  
30 GCACCGTGCTTCATGCGTCGGCTGACGCCTCCACCGTCGCCGAGCAGGTA

50

55

5

10

15

20

25

30

35

40

45

50

55

TCCGAAGCTGCCAGTCGCCGAAACGACTGGCAGGGAGTCCTCTACCTGTG  
GGGCCTCGACGCCGTCGTCGATGCTGGGGCATCGGCCGACGAAGTCAGCG  
AGGCTACCCGCCGTGCCACCGCACCCGTCCTTGGGCTGGTTCGATTCCCTG  
AGCGCTGCGCCCCATCCTCCTCGCTTCTGGGTGGTGACCCGCGGGGCATG  
5 CACGGTGGGCGGCGAGCCAGAGGCCTCTCTTTGCCAAGCGGCGTTGTGGG  
GCCTCGCGCGGTCGCGGGCGCTGGAGCACCCGCTGCCTGGGGTGGCCTC  
GTGGACCTGGATCCTCAGAAGAGCCCGACGGAGATCGAGCCCCTGGTGGC  
CGAGCTGCTTTCGCCGGACGCCGAGGATCAACTGGCGTTCGCGAGCGGTC  
GCAGGCACGCAGCACGCCTTGTAGCCGCCCGCCGGAGGGCGACGTCGCA  
10 CCGATATCGCTGTCCGCGGAGGGGAGCTACCTGGTGACGGGCGGGCTGGG  
TGGCCTTGGTCTGCTCGTGGCTCGGTGGCTGGTGGAGCGGGGAGCTCGAC  
ATCTGGTGCTCACCAGCCGGCACGGGCTGCCAGAGCGACAGGCGTCGGGC  
GGAGAGCAGCCGCCGAGGCCCGCGCGCATCGCAGCGGTGAGGGGGCT  
25 GGAAGCGCAGGGCGCGGGGTGACCGTGGCAGCGGTGGATGTGCGCGAGG  
CCGATCCCATGACGGCGCTGCTGGCCGCCATCGAGCCCCCGTTGCGCGGG  
GTGGTGACGCGCCCGGCGTCTTCCCGTGCGTCACCTGGCGGAGACGGA  
30 CGAGGCCCTGCTGGAGTCGGTGCTCCGTCCCAAGGTGGCCGGGAGCTGGC  
TGCTGCACCGGCTGCTGCGCGACCGGCCTCTCGACCTGTTCGTGCTGTTC  
TCGTGCGGCGCGGCGGTGTGGGGTGGCAAAGGCCAAGGCGCATACGCCGC  
20 GGCCAAATGCGTTCCCTCGACGGGCTC3CGCACCATCGCCGCGCGCACTCGC  
TGCCGGCGTTGAGCCTCGCCTGGGGCTTATGGGCCGAGGGAGGCATGGTT  
GATGCAAAGGCTCATGCAGTCTGAGCGACATCGGGGTCTGCCCATGGC  
CACGGGGCCGGCCTTGTCCGGCGCTGGAGCGCCTGGTGAACACCAGCGCTG  
40 TCCAGCGTTCCGTACACGGATGGACTGGGCGCGCTTCGCGCCGGTCTAT  
25 GCCGCGCAGGGCGGCGCAACTTGCTTTCGGCTCTGGTCGCGGAGGACGA  
GCGCGCTGCGTCTCCCCGGTGCCGACGGCAAACCGGATCTGGCGCGGCC  
TGTCGGTTGCGGAGAGCCGCTCAGCCCTCTACGAGCTCGTTCGCGGCATC  
45 GTCGCCCGGGTGCTGGGCTTCTCCGACCCGGGCGCGCTCGACGTGGGCCG  
AGGCTTCGCCGAGCAGGGGCTCGACTCCCTGATGGCTCTGGAGATCCGTA  
30 ACCGCCTTCAGCGCGAGCTGGGCGAACGGCTGTGCGCGACTCTGGCCTTC

5

164

10

15

20

25

15

30

35

40

25

45

30

50

55

GACCACCCGACGGTGGAGCGGCTGCTGGCGCATCTCCTCACCACGTGCT  
GAAGCTGGAGGACCGGAGCGACACCCGGCACATCCGGTCGGTGGCGGCGG  
ATGACGACATCGCCATCGTCGGTGCCGCTGCCGGTCCCAGGTGGGGAT  
GAGGGCCTGGAGACATACTGGCGGCATCTGGCCGAGGGCATGGTGGTCAG  
5 CACCGAGGTGCCAGCCGACCGGTGSCGCGCGGCGGACTGGTACGACCCCG  
15 ATCCGAGGTTCCGGGCCGGACCTATGTGGCCAAGGGTGCCTTCCTCCGC  
GATGTGCGCAGCTTGGATGCGGCGTTCTTCGCCATTTCCCCTCGTGAGGC  
GATGAGCCTGGACCCGCAACAGCGCTGTGCTGGAGGTGAGCTGGGAGG  
CGATCGAGCGCGCTGGCCAGGACCGATGGCGCTGCGCGAGAGCGCCACG  
20 10 GCGGTGTTCTGTGGGCATGATCGGGAGCGAGCACGCCGAGCGGGTGCAGGG  
CCTCGACGACGACGCGGCGTTGCTGTACGGCACCACCGCAACCTGCTCA  
GCGTCGCCGCTGGACGGCTGTGCTTCTTCCTGGGTCTGCACGGCCCCGACG  
25 ATGACGGTGGACACCGCCTGCTCGTCGTCGCTGGTGGCGTTGCACCTCGC  
CTGCCAGAGCCTGCGATTGGGCGAGTGGCAGCAGGCCCTGGCCGGCGGGT  
15 CCAGCGTGCTTTTGTGCGCCGCGGTCAATTCGTGCGGCGCTGCGCATGCGT  
TTGCTTTCGCCAGATGGGCGGTGCAAGACGTTCTCGGCCGCTGCAGACGG  
30 CTTTGC CGGGGCCGAGGGCTGCGCCGTGGTGGTGCTCAAGCGGCTCCGTG  
ACGCGCAGCGCGACCGCGACCCCATCTTGGCGGTGGTCAGGAGCACGGCG  
ATCAACCACGATGGCCCCGAGCAGCGGGCTCACGGTGGCCAGCGTCTCTGC  
20 CCAGCAGGCGTTGCTACGCCAGGCGCTGGCGCAAGCGGGCGTGGCGCCGG  
35 CCGAGGTGCAATTCGTGGAGTGCCACGGGACGGGACAGCGCTGGGTGAC  
CCGATCGAGGTGCAGGCGCTGGGCGCGGTGTACGGGCGGGGCGCCCCGC  
GGAGCGGCCGCTCTGGCTGGGCGCTGTCAAGGCCAACCTCGGCCACCTGG  
40 AGGCCGCGGGCGGCTTGGCCGGCGTGCTCAAGGTGCTCTTGGCGCTGGAG  
25 CACGAGCAGATTCCGGCTCAACCAGAGCTCGACGAGCTCAACCCGCACAT  
CCCGTGGGCAGAGCTGCCAGTGGCGGTGTCCGAGGGCGGTCCCCTGGC  
CGCGCGGCGCGCCCCGCGTCGTGACGGCGTGAGCGCTTTCGGCCTGAGC  
45 GGGACCAACGCGCATGTGGTGTGAGGAGGCGCCGGCGGTGGAGCCTGT  
GGCCGCGGCCCCGAGCGCGCAGCGAGCTGTTCTGCTCTGTCGGCGAAGA  
30 GCGCGCGGCGCTGGATGCGCAGCGAGCCGGCTGCGGGACCACCTGGAG

5

165

10

5

15

20

10

25

15

30

35

40

25

45

50

55

AAGCATGTCCAGCTTGGCCTCGGCGATGTGGCGTTCAGCCTGGCGACGAC  
GCGCAGCGCGATGGAGCACCGGCTGGCGGTGGCCGCGAGCTCGCGCGAGG  
CGCTGCGAGGGGCGCTTTCGGCCGCGAGCGAGGGGCACACGCCGCCGGGA  
GCCGTGCGTGGGCGGCCCTCGGGCGGCAGCGCGCCGAAGGTGGTCTTCGT  
GTTTCCCGGCCAGGGCTCGCAGTGGGTGGGCATGGGCCGAAAGCTCATGG  
CCGAAGAGCCGGTCTTCCGGGCGGCGCTGGAGGGTTGCGACCGGGCCATC  
GAGGCGGAAGCGGGCTGGTCTGCTCGGGGAGCTCTCCGCCGACGAGGC  
CGCCTCGCAGCTCGGGCGCATCGACGTGGTTAGCCGGTGCTGTTCCGCA  
TGGAAGTAGCGCTTCTGCGCTGTGGCGGTCTGCGGAGTGGAGCCGAA  
GCGGTGGTGGGCCACAGCATGGGCGAGGTTGCGGCGGCGACGTGGCCGG  
CGCGCTGTGCTCGAGGACGCGGTGGCGATCATCTGCCGGCGCAGCCGGC  
TGCTGCGGCGGATCAGCGGTGAGGGGAGATGGCGCTGGTCTGAGCTGTCTG  
CTGGAGGAGGCCGAGGCGGCGCTGCGTGCCATGAGGGTCTGGCTGAGCGT  
GGCGGTGAGCAACAGCCCGCGCTCGACCGTGCTCGCCGGCGAGCCGGCGG  
CGCTCTCGGAGGTGCTGGCGGCGCTGACGGCCAAGGGGTGTTCTGGCGG  
CAGGTGAAGGTGGACGTGCCAGCCATAGCCCGCAGGTGACCCGCTGCG  
CGAAGAGCTGATCGCGGCGCTGGGAGCGATCCGGCCGCGAGCGGCTGCGG  
TGCCGATGCGCTCGACGGTGACGGGCGGGGTGATCGCGGTTCCGGAGCTC  
GGTGCGAGCTACTGGGCGGACAACCTTCGGCAGCCGGTGCGCTTCGCTGC  
GGCGGCGCAAGCGCTGCTGGAGGGTGGCCCCGCGCTGTTTCATCGAGATGA  
GCCCCACCCGATCCTGGTGCCGCCCTGGACGAGATCCAGACGGCGGCC  
GAGCAAGGGGCGCTGCGGTGGGCTCGCTGCGGCGAGGGCAGGACGAGCG  
CGCGACGCTGCTGGAGGCGCTGGGGACGCTGTGGGCGTCCGGCTATCCGG  
TGAGCTGGGCTCGGCTGTTCCCGCGGGCGGAGGCGGTTCCGCTGCCG  
ACCTATCCCTGGCAGCACGAGCGGTACTGGATCGAGGACAGCGTGATGG  
GTCAAGCCCTCGCTGCGGCTTCGGCAGCTTCGCAACGGCGCCACGGACC  
ATCCGCTGCTCGGGGCTCCATTGCTCGTCTCGGCGGACCCGGAGCTCAC  
TTGTGGGAGCAAGCGCTGAGCGACGAGAGGCTATCCTACCTTTCGGAACA  
TAGGGTCCATGGCGAAGCCGTGTTGCCAGCGGCGGTATGTAGAGATGG  
CGCTCGCCCGGCGTAGATCTCTATGGCACGCGACGCTGGTGCTGGAG

5

10

15

20

25

30

35

40

45

50

55

CAGCTGGCGCTCGAGCGAGCCCTCGCCGTGCCCTCCGAAGCGGACGCAT  
CGTGCAAGTGGCCCTCAGCGAAGAAGGTCCCGGTGCGGCCTCATTCCAGG  
TATCGAGTCGTGAGGAGGCAGGTAGGAGCTGGGTGCGGCACGCCACGGGG  
CACGTGTGTAGCGGCCAGAGCTCAGCGGTGGGAGCGTTGAAGGAAGCTCC  
GTGGGAGATTCAACGGCGATGTCCGAGCGTCCTGTCGTGCGAGGCGCTCT  
ATCCGCTGCTCAACGAGCACGCCCTCGACTATGGTCCCTGCTTCCAGGGC  
GTGGAGCACGTGTGGCTCGGCACGGGGGAGGTGCTCGGCCGGGTACGCTT  
GCCAGGAGACATGGCATCCTCAAGTGGCGCCTACCGGATTCATCCCGCCT  
TGTTGGATGCATGTTTTAGGTGCTGACAGCGCTGCTCACCACGCCGGAA  
TCCATCGAGATTCGGAGGCGGCTGACGGATCTCCACGAACCGGATCTCCC  
GCGGTCCAGGGCTCCGCTGAATCAAGCGGTGAGTGACACCTGGCTGTGGG  
ACGCCGCGCTGGACGGTGGACGGCGCCAGAGCGCGAGCGTGCCCGTCGAC  
CTGGTGCTCGGCAGCTTCCATGCGAAGTGGGAGGTATGGAGCGCCTCGC  
GCAGGCGTACATCATCGGCACTCTCCGCATATGGAACGTCTTCTGCGCTG  
CTGGAGAGCGTCACACGATAGACGAGTTGCTCGTCAGGCTTCAAATCTCT  
GTCGTCTACAGGAAGGTCAATCAAGCGATGGATGGAACACCTTGTCGCGAT  
CGGCATCCTTGTAGGGGACGGAGAGCATTTTGTGAGCTCTCAGCCGCTGC  
CGGAGCCTGATTTGGCGGCGGTGCTCGAGGAGGCCGGAGGGTGTTGCGC  
GACCTCCCAGTCCCTATTTGAGTGGTGCAAGTTTGCCGGGGAACGGCTCGC  
GGACGTATTGACCGGTAAGACGCTCGCGCTCGAGATCCTCTTCCCTGGTG  
GCTCGTTTCGATATGGCGGAGCGAATCTATCGAGATTCGCCCATCGCCCGT  
TACTCGAACGGCATCGTGCGCGGTGTCGTGAGTCGGCGGCGCGGTGGT  
AGCACCGTCGGGAATGTTTACGATCCTTGAGATCGGAGCAGGGACGGGCG  
CGACCACCGCCCGCTCCTCCCGGTGTTGCTGCCTGACCGGACGGAGTAC  
CATTTACCGATGTTTCTCCGCTCTTCTTGCTCGCGCGGAGCAAAGATT  
TCGAGATTATCCATTCTGAAGTATGGCATTCTGGATGTCGACCAGGAGC  
CAGCTGGCCAGGGATACGCACATCAGAGGTTTGACGTCATCGTCGCGGCC  
AATGTCATCCATGCGACCCGCGATATAAGAGCCACGGCGAAGCGTCTCCT  
GTCGTTGCTCGCGCCCGGAGGCCTTCTGGTGCTGGTCGAGGGCACAGGGC  
ATCCGATCTGGTTCGATATCACCACGGGATTGATTGAGGGGTGCGAGAAG

5

10

15

20

25

30

35

40

45

50

55

TACGAAGATGATCTTCGTATCGACCATCCGCTCCTGCCTGCTCGGACCTG  
GTGTGACGTCCTGCGCCGGGTAGGCTTTGCGGACGCCGTGAGTCTGCCAG  
GCGACGGATCTCCGGCGGGGATCCTCGGACAGCACGTGATCCTCTCGCGC  
GCGCCGGGCATAGCAGGAGCCGCTTGTGACAGCTCCGGTGAGTCGGCGAC  
CGAATCGCCGGCCGCGCGTGCAGTACGGCAGGAATGGGCCGATGGCTCCG  
CTGACGTCGTCCATCGGATGGCGTTGGAGAGGATGTACTTCCACCGCCGG  
CCGGGCCGGCAGGTTTGGGTCCACGGTCGATTGCGTACCGGTGGAGGCGC  
GTTACGAAGGCGCTCGCTGGAGATCTGCTCCTGTTTGAAGACACCGGGC  
AGGTCGTGGCAGAGGTTTCAAGGGCTCCGCCTGCCGACGCTCGAGGCTTCT  
GCTTTCGCGCCGCGGGACCCGCGGGAAGAGTGTTGTACGCTTTGGAATG  
GCAGCGCAAAGACCCTATACCAGAGGCTCCGGCAGCCGCGTCTTCTTCCT  
CCGCGGGGGCTTGGCTCGTGCTGATGGACCAGGGCGGGACAGGCGCTGCG  
CTCGTATCGCTGCTGGAAGGGCGAGGCGAGGCGTGCCTGCGCGTCATCGC  
GGGTACGGCATAACGCTGCCTCGCGCCGGGGCTGTATCAAGTCGATCCGG  
CGCAGCCAGATGGCTTTCATACCCTGCTCCGCGATGCATTGGGCGAGGAC  
CGGATTTGTGCGCGGTTAGTGATATGTGGAGCCTTGATGCGACGGCAGC  
AGGGGAGAGGGCGACAGCGGAGTCGCTTCAGGCCGATCAACTCCTGGGGA  
GCCTGAGCGCGCTTTCTCTGGTGCAGGCGCTGGTGCGCCGAGGTGGCGC  
AACATGCCGCGGCTTTGGCTCTTGACCCGCGCCGTGCATGCGGTGGGCGC  
GGAGGACGCGAGCGCCTCGGTGGCCGAGGCGCCGGTGTGGGGCCTCGGTC  
GGACGCTCGCGCTCGAGCATCCAGAGCTGCGGTGCACGCTCGTGGACGTG  
AACC CGGCGCCGTCTCCAGAGGACGAGCCGCACTGGCGGTGGAGCTCGG  
GGCAGCGACAGAGAGGACCAGGTGCGATTGCGCTCGGATGGCCGCTACG  
TGGCGCGCCTCGTGCGGAGCTCCTTTCCGGCAAGCCTGCTACGGATTGC  
GGCATCCGGGCGGACGGCAGCTATGTGATCACCATGGCATGGGGAGAGT  
GGGGCTCTCGGTGCGCAATGGATGGTATGACAGGGGGCCCGCCATGTGG  
TGCTCGTGGATCGCGGCGGCGCTTCCGAGGCATCCCGGGATGCCCTCCGG  
TCCATGGCCGAGGCTGGCGCGGAGGTGCAGATCGTGGAGGCCGACGTGGC  
TCGGCGCGACGATGTCGCTCGGCTCCTCTCGAAGATCGAACCGTCCGATGC  
CGCCGCTTCCGGGGATCGTGTACGTTGGACGGGACCTTCCAGGGCGACTCC



5

10

15

20

25

30

35

40

45

50

55

TCGATGCTGGAGCTGGATGCCCCGTCGCTTCAAGGAGTGGATGTATCCCAA  
GGTGCTCGGAGCGTGGAACCTGCACGCGCTGACCAGGGATAGATCGCTGG  
ACTTCTTCGTCTGTATTCTCTCGGCGACCTCGCTTCTGGGCTTGCCAGGA  
CAGGGGAGCCGCGCCGCGGTGACGCCTTCTTGGACGCCATCGCGCATCA  
CCGGTGCAAGGTGGGCCTTACAGCGATGAGCATCAACTGGGGATTGCTCT  
CCGAAGCATCATCGCCGCGACCCCGAACGACGGCGGAGCACGGCTCGAA  
TACCGGGGATGGAAGGCCTCACGCTGGAGCAGGGAGCGGCGGCGCTCGG  
GCGCTTGCTCGCACGACCCAGGGCGCAGGTAGGGGTGATGCGGCTGAATC  
TGCGCCAGTGGTTGGAGTXXXXXXXXXXXXXXXXXXXXXXXXXXXXX  
XXXXTATGGTATAACTTATTGATTATAATACAGTATACAAAGGTCCCTTT  
TCAGSGACCCCTTCGTATGTTGTAGCTGATTTTATTTCTTCTTTCTTT  
TGGGTGCTTTTAATAGCCATATAATGATGTTTCATATAAAGTTAAAAGTTT  
CCCCATCTGTCTATCTATAGCGTCATGTTTTTCAGGATTTCTTAATTTCT  
GCAGCAGTTTAGATATATATTTTCGCAGGTAACCTTATTATTAATGAATTA  
TGGGATTGATACATTTCCCTTTGTCCAATCTAAATCTTTCTGATTTCAA  
TACCTCACTCTCATCTCTATTAATTATAAAACCGAGCTTTCCATACGGAC  
CTGTAAATATGATTGTAATTGTCTATACTCAGATGGACCTAATTTCTTCA  
AAATTTTATAGCATCAAAAACAACCTTGCTTGTTTTATAGTCTCCAATAC  
TCGTTTCCAAAAATCAGATTTGCCACCATTGGTGCCTATAATATCTCGTC  
TCTGAACCTGCGTTACCATTTGGATGGGACTTGATGTCTGTTAGGTGGGAT  
GCAAATACTATTCTTAGTGCGTCTAAACACCATTGTTCAAATTCAGTGGC  
ACCTTCATTTCCATATCGGTATCTGATCTAGATGAGTAGTTATTTGACCTA  
TTGTTTTATTTCTAATGGCTGAATTATCTGAAATAATATTAATATCATAT  
TCGTCAATTTATCTCCTCTGCTTCCCTCTGGAGCAAGTGCATTACGATTTAG  
ATTCAAACCAAGCCAGTAACATGGATGAATTAATAATTTCTCATTACTTT  
CAAAACCTTTATCTGGAGTTTCGCGCGTCATGGCAAAATGAATAGGATGAG  
GTGTTTTATCACGAATACCAACAAATCCTACACTATACAACTTTGGAG  
AATCCCACTTGCCCTTTAACAACCTGATTTTCAGACGTTATTTTAGGATCTC  
CATTTTCTTCGATTAATTCGAAAGATGCTTCTATTTTTTTAAGCACGTA  
TAAACTGTTAATTCAGGTTCAATGCTACGAAATGCACTAGTTATAACCTG

5

169

10

5

15

20

10

25

15

30

35

40

25

45

30

50

55

TATTGAAGGAAAGATCTTCTGATACTCTTTCCAGAGATCTTCAAGTCTGG  
CCATGGAAATTGACTTGGCTGCATATTCTAGGTCAGTGTATGATAGTT  
TCTCTATTCTCTCTGAATGCGGAAAAAAGCTTCATTCAACAATGATAG  
TAAATCCCTGGGCCGGTAAAGGGTAAATTGCAAACATCGCTTAAACCAT  
TCCTCCCTTTAAGATCATCCGCTGTGCATCTATCCCAAACCTCGTTGATCT  
TTCTCAATATCTAGCTTAAATGCTACTTTTCATTCTTTTAGCTGACAGCAT  
TAGGAGTTGTGCCCAGTCCCAATGCAACCTTATGACTTGACCCTCTATAT  
TTCTCGAGTAATCAGGATCTTCCTTTGATAGCGACCTAAATATATTATCC  
CTTAAAAAATTATTTGGACGAATGCATTTTGCTTTTGTATTTAATTCAAT  
AGATGCATATGCTAGACCTGCAATGATTCCAATTCCTATATTATCCGGTT  
CATACGCCTCATCTAGCTTATCCATTAATATGACAACCTTTCCTGTCTGAG  
CGTTCAAGAAGTGATACTATATTATTTTCTATTTCTGAGATATTCAAATT  
GAATTGAAGATCACCAATTGATTCTTCTGGGTTATTTTCATCTAAATACT  
CCTTTGCGACAAGCCTACACTTTCTTAAATGTCACCTTGTGCAGAATTC  
CATTTTTTCAAATGTTTCATTCAACAATGTTTCTGATGATATTTGAGATGA  
CAATTTGTAATGAGATGATATATATGATGCTATCTCCATTAGCATAGCGT  
ATCGCCATAATAGTCTTGTTGCTGCCCTTGCTAAATTAATGATCCTGTA  
AATGGTTTTCAACATTGATCTGAAACCAATAATTGAGAATCGTCTGGTGA  
GAAACTCAGGATTAATATTTTTTTGTCTTTCTTCCAATGCTCATTTAGCT  
GAATAAATAAAGCACTTTTACCTGTCCCTCGTCTACCAACAACAATGGTC  
CTGTCTACGTTTCAATTAGAGTCCTAAAGTCAGCAGTTTCAATGAAAGC  
ATTACTCAACATCTTTTTATCATTTTCTGCTGTCTATCACCACCGGAT  
TAGACTTTGAAGTAATATTCAATTCCATATTCAACCTTTTATGTTAGTTG  
CTTTCATTTATTACTTTATATACTGTTGAACGAGCAATATTCATTGTTTT  
TGATATATGTGAGGCACCTAACCCTGTTGCCACATATTTAATACTGCAT  
CTCTATCTATTTTCTTTTTCTACCAAAAACAACCTCTTTTGCCA

SEQ ID No 82 (>Contig57)

TCATCTATTGTATAGTTTGTATATTGATATGATATAATTATAACATATAA  
ACAGTAACTTTCTCTACGTAGATCGAGGAGAAGACTCAATTTGTTGACA

5

170

10

TCXXXXXXXXXXXXXXXXXXXXXXXXXXXXACTACTCGCATACCGTTG  
CGCAACAGCGGCGCAGGAGCAGGACGCATACGACATCACCGGCAATACG  
CTCAGCGTCGCCGACCGACGGTTGTCTTATACGCTAGGGCTGCAGGGACC  
CTGCCTGACCGTCGACACGGTCTGCTCGTCGTCGCTCGTGGCCATCCACC  
5 TTGCCTGCCGAGCCTGCGCGCTCGCGAGAGCGATCTCGCGCTGGCGGGA  
15 GGCGTCAACATGCTCCTTTTCGTCCAAGACGATGATAATGCTGGGGCGCAT  
CCAGGCGCTGTCGCCGATGGCCACTGCCGGACATTCGACGCCTCGGCCA  
ACGGGTTTCGTCCGTGGGGAGGGCTGCGGTATGGTCGTGCTCAAACGGCTC  
20 TCCGACGCCAGCGACACGCGATCGGATCTGGGCTCTGATCCGGGGTTC  
10 GGCCATGAATCAGGATGGCCGGTCGACAGGGTTGATGGCACCCAATGTGC  
TCGCTCAGGAGGCGCTCTTGCGCGAGGCGCTGCAGAGCGCTCGCGTCGAC  
GCCGGGGCCATCGGTTATGTGAGACCCACGGAACGGGGACCTCGCTCGG  
25 CGACCCGATCGAGGTCGAGGCGCTGCGTGCCGTGTTGGGGCCGGCGCGG  
CCGATGGGAGCCGCTGCGTGCTGGGCGCAGTGAAGACAAACCTCGGCCAC  
15 CTGGAGGGCGCTGCAGGCGTGGCGGGTTTGATCAAGGCGGCGCTGGCTCT  
GCACCACGAACTGATCCCGCGAAACCTCCATTTCCACACGCTCAATCCGC  
30 GGATCCGGATCGAGGGGACCGCGCTCGCGCTGGCGACGGAGCCGGTGCCG  
TGGCCGCGGGCGGGCCGACCGCGCTTCGCGGGGGTGAGCGGTTCCGGCCT  
CAGCGGCACCAACGTCCATGTCGTGCTGGAGGAGGCGCCGGCCACGGTGC  
35 20 TCGCACCGGCGACGCCGGGGCGCTCAGCGGAGCTTTTGGTGCTGTCCGCG  
AAGAGCGCCGCGCGCTGGACGCACAGGCGGCGCGGCTCTCAGCGCACAT  
CGCCCGGTACCCGGAGCAGGGTCTCGGAGACGTGCGGTTACAGCCTGGTAT  
CGACGCGTAGCCCGATGGAGCACCGGCTCGCGGTGGCGGCGACCTCGCGC  
40 GAGGCGCTGCGAAGCGCGCTGGAGGTTGCGGCGCAGGGGCAGACCCCGGC  
25 AGGCGCGGCGCGCGGAGGGCCGCTTCCTCGCCCGCAAGCTCGCCTTCC  
TGTTCCCGGGCAGGCGCGCAGGTGCCGGGCATGGGCCGTGGGTTGTGG  
GAGGCGTGCCCGGCGTTCCGCGAGACCTTCGACCGGTGCGTCACGCTCTT  
45 CGACCGGGAGCTCCATCAGCCGCTCTGCGAGGTGATGTGGGCCGAGCCGG  
GCAGCAGCAGGTGCTCGTTGCTGGACCAGACGGCGTTACCCAGCCGGCG  
30 CTCTTTGCGCTGGAGTACGCGCTGCGCGCGCTCTTCGGTTCGTGGGGCGT

50

55

5

10

15

20

25

30

35

40

45

50

55

GGAGCCGGAGCTCGTCGCTGGCCATAGCCTCGGCGAGCTGGTGGCCGCCT  
GCGTGGCGGGTGTGTCTCCCTCGAGGACGCCGTGCGCTTGGTGGTCGCG  
CGCGGCCGGTTGATGCAGGCGCTGCCGCCGGCGGCGCGATGGTATCGAT  
CGCCGCGCCGGAGGCCGACGTGGCTGCCGCCGTGGCGCCGCACGCAGCGT  
5 TGGTGTGATCGCGGCAGTCAATGGGCCGGAGCAGGTGGTGTATCGCGGC  
GCCGAGAAATTCTGTGCAGCAGATCGCGCGGCGTTGCGGCGCGGGGGGC  
GCGAACCAAACCGCTGCATGTCTCGCACGCGTTCCACTCGCCGCTCATGG  
ATCCGATGCTGGAGGCGTTCCGGCGGGTGAAGTGAAGTGAAGTGAAGTGAAG  
CGGCCTTCGATCGCGCTGGTGAGCAACCTGAGCGGGAAGCCCTGCACCGA  
10 TGAGGTGAGCGCGCCGGTTACTGGTGCGTCACGCGCGAGAGGCGGTGC  
GCTTCGCGGACGGAGTGAAGGCGCTGCACGCGGCCGGTGCGGGCCTCTTC  
GTCGAGGTGGGGCCGAAGCCGACGCTGCTCGGCCTTGTGCCGGCCTGCCT  
GCCGGATGCCAGGCCCGTGCTGCTCCAGCGTCGCGCGCCGGCGGTGACG  
25 AGGCTGCGAGCGCGCTAGAGGCGCTGGGTGGGTTCGGGTGCTCGGTGGA  
TCGGTCACCTGGTCCGGTGTCTTCCTTCGGGCGGACGGCGGTACCGCT  
GCCAACCTATCCCTGGCAGCGCGAGCGTTACTGGATCGAAGCGCCGGTCG  
ATCGTGAGGCGGACGGCACCGGCCGTGCTCGGGCGGGGGCCACCCCTT  
CTGGGTGAAGTCTTTTCCGTGTCGACCCATGCCGCTGCGCCTGTGGGA  
GACGACGCTGGACCGAAAGCGGCTGCCGTGGCTCGGCGAGCACCGGGCGC  
20 AGGGGGAGGTGCTGTTTCTGGCGCCGGGTACCTGGAGATGGCGCTGTCG  
TCGGGGGCCGAGATCTTGGGCGATGGACCGATCCAGGTACCGATGTGGT  
GCTCATCGAGACGCTGACCTTCGCGGGCGATACGGCGGTACCGGTCCAGG  
TGGTGACGACCGAGGAGCGACCGGGACGGCTGCGGTTCCAGGTAGCGAGT  
40 CGGGAGCCGGGGGAACGTCGCGCGCCCTTCGGATCCACGCCCGCGGCGT  
GCTGCGCCGGATCGGGCGCGTCGAGACCCCGCGAGGTGGAACCTCGCCG  
CCCTGCGCGCCCGGCTTCATGCCGCCGTGCCCGCTGCGGCTATCTATGGT  
GCGCTCGCCGAGATGGGGCTTCAATACGGCCCGGCGTTGCGGGGGCTCGC  
45 CGAGCTGTGGCGGGGTGAGGGCGAAGCGCTGGGCGAGGTGAGACTGCCTG  
AGGCCGCCGCTCCGCGACAGCCTACCAGCTGCATCCGGTGTGCTGGAC  
30 GCGTGCGTCCAAATGATTGTTGGCGCGTTCCGCCATCGCGATGAGGCGAC

5

172

10

5

15

20

10

25

15

30

35

20

40

25

45

30

50

55

GCCGTGGGCGCCSGTGGAGGTGGGCTCGGTGCGGCTGTTCCAGCGGTCTC  
CTGGGGAGCTATGGTGCCATGCGCGCGTCGTGAGCGATGGTCAACAGGCC  
TCCAGCCGGTGGAGCGCCGACTTTGAGTTGATGGACGGTACGGGCGCGGT  
GGTCGCCGAGATCTCCCGCTGCTGGTGGAGCGGCTTGCAGCGGTGTAC  
GCCGGCGGACGACGAGCACTGGTTCCCTGGAGCTGGATTGGGAGCCCGCG  
GCGCTCGGTGGGCCCAAGATCACAGCCGGCCGGTGGCTGCTGCTCGGCCA  
GGGTGGTGGGCTCGGGCGCTCGTTGTGCTCGGCGCTGAAGGCCCGCGGCC  
ATGTCGTGCTCCACGCCCGGGGGACGACAGGCACTGCAGGAATGCGC  
GCGCTCCTGGCCAACCGTTTCGACGGCCAGGCCCGACGGCCGTGGTGCA  
CCTCAGCAGCCTCGACGGGGCGGCCAGCTCGGCCCGGGGCTCGGGGCGC  
AGGGCGGCTCGACGCGCCCGGAGCCAGATGTCGATGCCGATGCCCTC  
GAATCGGCGCTGATGCGTGGTTGCGACAGCGTGCTCTCCCTGGTGCAAGC  
GCTGGTCGGCATGGACCTCCGAAACGCGCCGCGGCTGTGGCTCTTGACCC  
GCGGGGCTCAGGCGGCCCGCCGGCGATGTCTCCGTGGTGCAAGCGCCG  
CTGTTGGGGCTGGGCCGACCATCGCCTTGAGCACGCCGAGCTGCGCTG  
TATCAGCGTCGACCTCGATCCAGCCGAGCCTGAAGGGGAAGCCGATGCTT  
TGCTGGCCGAGCTACTTGACAGATGATGCCGAGGAGGAGGTGCGCTGCGC  
GGTGGCGACCGGCTCGTTGCGCGGCTCGTCCACCGGCTGCCCCGACGCTCA  
GCGCCGGGAGAAGGTCGAGCCCGCCGGTGACAGGCCGTTCCGGCTAGAGA  
TCGATGAACCCGGCGCGCTGGACCAACTGGTGCTCCGAGCCACGGGGCGG  
CGCCTCCTGGTCCGGGCGAGGTCGAGATCTCCGTGCAAGCGGCGGGGCT  
CGACTCCATCGACATCCAGCTGGCGTTGGGCGTTGCTCCCAATGATCTGC  
CTGGAGAAGAAATCGAGCCGTTGGTGCTCGGAAGCGAGTGCGCCGGGCGC  
ATCSTCGCTGTGGGCGAGGGCGTGAACGGCCTTGTGGTGGGCCAGCCGGT  
GATCGCCCTTGCGGCGGGAGTATTTGCTACCCATGTCACCACGTGGCCA  
CGCTGGTGTGCTCGGCCTCTGGGGCTCTCGGCGACCGAGGCGGCCGCG  
ATGCCCCCTCGCGTATTTGACGGCCTGGTACGCCCTCGACAAGGTCCGCCA  
CCTGCAGGCGGGGAGCGGGTGCTGATCCATGCGGAGGCCGGTGGTGTCG  
GTCTTTGCGCGGTGCGATGGGCGCAGCGGTGGGCGCCGAGGTGTATGCG  
ACC3CCGACACGCCCCGAGAACCGTGCCTACCTGGAGTCGCTGGGCGTGCG

5

10

15

20

25

30

35

40

45

50

55

GTACGTGAGCGATTCCCGCTCGGGCCGGTTCGTACAGACGTGCATGCAT  
GGACGGACGGCGAGGGTGTGGACGTCGTGCTCGACTCGCTTTCGGGCGAG  
CGCATCGACAAGAGCCTCATGGTCCTGCGCGCCTGTGGTCGCCTTGTGAA  
GCTGGGCAGGCGCGACGACTGCGCCGACACGCAGCCTGGGCTGCCGCCGC  
5 TCCTACGGAATTTTTCCTTCTCGCAGGTGGACTTGCGGGGAATGATGCTC  
GATCAACCGGCGAGGATCCGTGCGCTCCTCGACGAGCTGTTGCGGTTGGT  
CGCAGCCGGTGCCATCAGCCCACTGGGGTTCGGGTTGCGCGTTGGCGGAT  
CCCTCACGCCACCGCCGGTCGAGACCTTCCCGATCTCTCGCGCAGCCGAG  
GCATTCCGGAGGATGGCGCAAGGACAGCATCTCGGAAGCTCGTGCTCAC  
10 GCTGGACGACCCGGAGGTGCGGATCCGCGCTCCGGCCGAATCCAGCGTCG  
CCGTCCGCGCGGACGGCACCTACCTTGTGACCGGCGGTCTGGGTGGCCTC  
GGTCTGCGCGTGGCCGGATGGCTGGCCGAGCGGGGCGCGGGCAACTGGT  
GCTGGTGGGCCGCTCCGGTGCGGCGAGCGCAGAGCAGCGAGCCCGGTGG  
25 CGGCGCTGGAGGCCACGGCGCGCGCTCACGGTGGCGAAAGCGGACGTC  
GCCGATCGGTCACAGATCGAGCGGGTCTCCGCGAGGTTACCGCCTCGGG  
GATGCCGCTGCGGGGTGTCGTGCATGCGGCAGGTCTCGTGGATGACGGGC  
TGCTGATGCAGCAGACTCCGGCGCGGTCCGACGGTGTGGGACCTAAG  
GTCCAGGGGGCCTTGCACTTGACACGCTGACACGCGAAGCGCCTCTTTC  
CTTCTTCGTGCTGTACGCTTCTGCAGCTGGGCTTTTCGGCTCGCCAGGCC  
20 AGGGCAACTATGCCGCGAGCCAACGCGTTCCTCGACGCCCTTTCGCATCAC  
CGAAGGGCGCAGGGCCTGCCGGCGCTGAGCATCGACTGGGGCATGTTTAC  
GGAGGTGGGGATGGCCGTTGCGCAAGAAAACCGTGGCGCGCGGCAGATCT  
CTCGCGGGATGCGGGGCATCACCCCGATGAGGGTCTGTGAGCTCTGGCG  
CGCTTGCTCGAGGGTGATCGCGTGACAGCGGGGTGATACCGATCACTCC  
25 GCGGCAGTGGGTGGAGTTCTACCCGGCAACAGCGGCCTCACGGAGGTTGT  
CGCGGTGGTGACCACGACGCGCGCGTCTGATCGGACCGCCGGGGAT  
CGGGACCTGCTCGAACAGCTTGCGTTCGGCTGAGCCGAGCGCGGGCGGG  
GCTGCTGCAGGACGTCGTGCGCGTGAGGTCTCGCATGTGCTGCGTCTCC  
45 CTGAAGACAAGATCGAGGTGGATGCCCCGCTCTCGAGCATGGGCATGGAC  
30 TCGCTGATGAGCCTGGAGCTGCGCAACCGCATCGAGGCTGCGCTGGGCGT

5

10

15

20

25

30

35

40

45

50

55

CGCCGCGCCTGCAGCCTTGGGGTGGACGTACCCAACGGTAGCAGCGATAA  
CGCGCTGGCTGCTCGACGACGCCCTCGTCGTCCGGCTTGGCGGCGGGTCC  
GACACGGACGAATCGACGGCGAGCGCCGGTTCGTTTCGTCCACGTCCTCCG  
CTTTCGTCTGTCTGTCAGCCGCGGGCTCGTCTCTTCTGTTTTCACGGTT  
CTGGCGGCTCGCCCGAGGGCTTCCGTTCCCTGGTCCGAGAAGTCTGAGTGG  
AGCGATCTGGAATCGTGGCCATGTGGCACGATCGCAGCCTCGCCTCCGA  
GGACGCGCCTGGTAAGAAGTACGTCCAAGAGGCGGCCTCGCTGATTACAGC  
ACTATGCAGACGCACCGTTTTCGCTTAGTAGGGTTCAGCCTGGGTGTCCGG  
TTCGTCAATGGGGACAGCCGTGGAGCTCGCCAGTCGTTCCGGCGCACCGGC  
TCCGCTGGCCGTCTTACGTTGGGCGGCAGCTTGATCTCTTCTTACAGAGA  
TCACCCCGGAGATGGAGACCGATATAATAGCCAAGCTCTTCTTCCGAAAT  
GCCGCGGGTTTCGTGCGATCCACCCAACAAGTCCAGGCCGATGCTCGCGC  
AGACAAAGTCATCACAGACACCATGGTGGCTCCGGCCCCCGGGACTCGA  
AGGAGCCGCGCGTGAAGATCGCGGTCCCTATCGTCGCCATCGCCGGCTCG  
GACGATGTGATCGTGCCTCCGAGCGACGTTTCCAGGATCTACAATCTCGCAC  
CACGGAGCGCTTCTATATGCATCTCCTTCCCGGAGATCACGAATTTCTCG  
TCGATCGAGGGCGCGAGATCATGCACATCGTCGACTCGCATCTCAATCCG  
CTGCTCGCCGCGAGGACGACGTCGTCAGGCCCCGCGTTCGAGGCAAAATG  
ATGGCAGCCTCCCTCGGGCGCGGAGATGGTTGGGAGCAGCGTGGGCGCT  
GGCGGCCGGCGGCAGGCCGCGGAGGCGCATGAGCCTTCTTGACGTTTGC  
AGTATAGGAGATTTTATGACACAGGAGCAAGCGAATCAGAGTGAGACCAA  
GCCTGCTTTTCGACTTCAAGCCGTTTCGCGCCTGGGTACGCGGAGGACCCGT  
TCCCCGCGATCGAGCGCCTGAGAGAGGCAACCCCATCTTCTACTGGGAT  
GAAGGCCGCTCCTGGGTCTTCACCCGATACACGACGTGTCGGCGGTGTT  
CCGCGACGAACGCTTCGCGGTCACTCGAGAAGAGTGGGAATCGAGCGCGG  
AGTACTCGTCGGCCATTCCCGAGCTCAGCGATATGAAGAAGTACGGATTG  
TTCGGGCTGCCGCCGAGGATCACGCTCGGGTCCGCAAGCTCGTCAACCC  
GTCGTTTACGTACGCGCCATCGACCTGCTGCGCGCCGAAATACAGCGCA  
CCGTCGACCAGCTGCTCGATGCTCGCTCCGGACAAGAGGAGTTCGACGTT  
GTGCGGGATTACGCGGAGGGAATCCCGATGCGCGCGATCAGCGCTCTGTT

5

10

15

20

25

15

30

35

40

25

45

30

50

55

GAAGGTTCCGGCCGAGTGTGACGAGAAGTTCCGTCGCTTCGGCTCGGCGA  
CTGCGCGCGCGCTCGGCGTGGGTTTGGTGCCCCAGGTCGATGAGGAGACC  
AAGACCCTGCTCGCGTCCGTACCGAGGGGCTCGCGCTGCTCCATGACGT  
CCTCGATGAGCGGCGCAGGAACCCGCTCGAAAATGACGTCTTGACGATGC  
TGCTTCAGGCCGAGGCCGACGGCAGCAGGCTGAGCACGAAGGAGCTGGTC  
GCGCTCGTGGGTGCGATTATCGCTGCTGGCACCGATAACCACGATCTACCT  
TATCGCGTTCGCTGTGCTCAACCTGCTGCGGTGCCCCGAGGCGCTCGAGC  
TGGTGAAGGCCGAGCCCGGGCTCATGAGGAACGCGCTCGATGAGGTGCTC  
CGCTTCGACAATATCCTCAGAATAGGAACTGTGCGTTTCGCCAGGCAGGA  
CCTGGAGTACTGCGGGGCATCGATCAAGAAAGGGAGATGGTCTTTCTCC  
TGATCCCAGCGCCCTGAGAGATGGGACTGTATTCTCCAGGCCAGACGTG  
TTTGATGTGCGACGGGACACGGGCGCGAGCCTCGCGTACGGTAGAGGCCC  
CCATGTCTGCCCCGGGTGTCCCTTGCTCGCCTCGAGGCGGAGATCGCCG  
TGGGCACCATCTCCGTAGGTTCCCCGAGATGAAGCTGAAAGAACTCCC  
GTGTTTGGATAACCACCCCGGTTCCGGAACATCGAATCACTCAACGTCAT  
CTTGAAGCCCTCCAAAGCTGGATAGCTCGCGGGGGTATCGCTTCCCGAAC  
CTCATTCCCTCATGATACAGCTCGCGCGCGGGTGCTGTCTGCCGCGGGTG  
CGATTGATCCAGCGACAAGCCCATTTGTCAGCGCGCGAAGATCGAATCC  
ACGGCCCCGAGAAGAGCCCGTCCGGGTGACGTCGGAAGAAGTGCCGGGCG  
CCGCCCTGGGAGCGCAAAGCTCGCTCGTTCGCGCTCAGCACGCCGCTCGT  
CATGTCCGGCCCTGCACCCGCGCCGAGGAGCCGCCGCCCTGATGCACGG  
CCTCACCGAGCGGCAGGTTCTGCTCTCGCTCGTCCCTCGCGCTCGTCC  
TCCTGACCGCGCGCCCTTCGGCGAGCTCGCGCGGCGGCTGCGCCAGCCC  
GAGGTGCTCGGCGAGCTCTTCGGCGGCGTGGTGCTGGGCCCGTCCGTCGT  
CGGCGCGCTCGCTCCTGGGTTCCATCGAGTCTCTTCCAGGATCCGGCGG  
TCGGGGTCTGCTCTCCGGCATCTCTGGATAGGCGCGCTCGTCTCTGCTG  
CTCATGGCGGGTATCGAGGTCGATGTGAGCATCCTGCGCAAGGAGGCGCG  
CCCCGGGCGCTCTCGGCGCTCGGCGCGATCGCGCCCCCGCTGCGCACGC  
CGGGGCGGCTGGTGACGCGCATGCAGGGCGGTTACGTGGGATCTCGAC  
GTCTCGCCGCGACGCTCTGCGCAAGCCTGAGCCTCGGCGCCTGCTCGTAC



5

176

10

ACCTCGCCGGTGCTCGCTCCGCCCGCGGACATCCGGCCGCCCGCCGCGGC  
CCAGCTCGAGCCGGACTCGCCGGATGACGAGGCCGACGAGGCCGACGAGG  
CGCTCCGCCCGTTCCGCGACGCGATCGCCGCGTACTCGGAGGCCGTTTCGG  
TGGGCGGAGGCGGCGCAGCGCCGCGGCTGGAGAGCCTCGTGCGGCTCGC  
5 GATCGTGCGGCTGGGCAAGGCGCTCGACAAGGTCCCTTTTCGCGCACACGA  
15 CGGCCGCGCTCTCCAGATCGCCGGCAGACTCCAGAACGATGCGGTCTGG  
TTCGATGTCGCCGCCCGGTACGCGAGCTTCCGCGCGGCGACGGAGCACGC  
GCTCCGCGACGCGGCGTCCGCCATGGAGGCGCTCGCGGCCGCCCGTACC  
GCGGATCGAGCCGCGTGTCCGCTGCCGTAGGGGAGTTTCGGGGGGAGGCG  
20 GCGCGCCTTACCCCGCGGACCGTGTACCCGCGTCCGACCAGCAGATCCT  
10 GACCGCGCTGCGCGCAGCCGAGCGGGCGCTCATCGCGCTCTACACTGCGT  
TCGCCCCGTGAGGAGTGAGCCTCTCTCGGGCGCAGCCGAGCGGCGGCGTGC  
25 CGGTGGTTCCTCTTCGCAACCATGACCGGAGCCGCGCTCGGTCCGCGCA  
GCGGCTAGCGCGCGTCCGCGCAGAGATCGCTGGAGCGACAGGCGACGACC  
15 CGCCCCGAGGGTGTGCAACGGATTGCCGCGAGCCCTCATTGCGGATCCCCTC  
CAGACACTCGTTTCAGCTGCTTGGCGTCGATGCCGCTGGGCACTCGCCGA  
30 AGGTCACTCGTCGCGCCACTCGGATCGGATCTTGTTTCGAGCACGCGTCC  
TTGCTCGAATACTCCCGGTCTTGTCCGATGTTGTTGCACCGCGCCTCGCG  
GTCGCACCGCGCCGCCACGATGCTATCGACGGCGCTGCCGACTGGCACCG  
20 GCGCCTCGCCCTGCGCGCCACCCGGGGTTTGCGCCTCCCGCCTGACCGC  
35 TTTTCGCCCGCCGACGCCGCGAGCAGGCTCATTCCCGACACCGAGATCAG  
GCCCACGACCAGCTTCCAGCAATCTTTTGCATGGCTTCCCTCCCTCAC  
GACACGTCACATCAGAGACTCTCCGCTCGGCTCGTCGGTTCGACAGCCGG  
40 CGACGCCACGAGCAGAACCGTCCCCGACCAGAACAGCCGCATGCGGGTT  
25 TCTCGAACATGCCCCGACATCCTTGCGACTAGCGTGCCCTCCGCTCGTGC  
CGAGATCGGCTGTCTGTGCGACGGCAATATCCTGCGATCGGCCGGGCAG  
GAGGTACCGACACGGGCGCCGGGCGGGAGGTGCCGCCACGGGCTCGAAAT  
45 GTGCTGCGGCAGGCGCCTCCATGCCCGCAGCCGGGAACGCGGCGCCCGGC  
CAGCCTCGGGGTGACGCCGCAAACGGGAGATGCTCCCGGAGAGGCGCCGG  
30 GCACAGCCGAGCGCCGTACCAACCGTGCGCACTCGTGAGCTCCAGCTCCT

50

55

5

10

15

20

25

30

35

40

45

50

55

CGGCATAGAAGAGACCGTCACTCCCGGTCCGTGTAGGCGATCGTGCTGAT  
CAGCGCGTTCTCCGCCTGACGCGAGTCGAGCCGGGTATGCTGCACGACAA  
TGGGAACGTCCGATTTCGATCACGCTGGCATACTCCGTATCGCGCGGGATC  
GGCTCGGGTTCGGTCAGATCGTTGAACCGGACGTGCCGGGTGCGCCTCGC  
TGGGACGGTCACCCGGTACGGCCCGGCGGGGTGCGGGTTCGCTGAAGTAGA  
CGGTGATGGCGACCTGCGCGTCCCGGTCCGACGCATTCAACAGGCAGGCC  
GTCTCATGGCTCGTCATCTGCGGCTCGGGTCCGTTGCTCCGGCCTGGGAT  
GTAGCCCTCTGCGATTGCCAGCGCGTCCGCCCCGATCGGCTTCTCCATAT  
GTCCTCCCTGCTGGCTCCTCTTTGGCTGCCCTCCCTCTGCTGTCCAGGAGC  
GACGGCCTCTTCTCCCGACGCGCTCGGGGATCCATGGCTGAGGATCCTCG  
CCGAGCGCTCCTTGCCGACCGGCGCGCCGAGCGCCGACGGGCTTTGAAAG  
CACGCGACCGGACACGTGATGCCGGCGCGACGAGGCCGCCCCGCGTCTGA  
TCCCGATCGTGACATCGCGACGTCCGCCGGCGCCTCTGCAGGCCGGCCTG  
AGCGTTGCGCGGTTCATGGTCGTCTCGCGTCAACGCCACCCGCCGATTCA  
CATCCACCGCGGCACGACGCTTGCTCAAACCGGGCGAGACGGCCGGGC  
GGCTGTGGTACCGGCCAGCCCCGACGCGAGGCCCGAGAGGGACAGTGGGT  
CCGCCGTGAAGCAGTGAGGCGATCGAGGTGGCAGATGAAACACGTTGACA  
CGGGCCGACGAGTCGGCCCGCGGATAGGGCTCACGCTCGGTCTCCTCGCG  
AGCATGGCGCTCGCCGGCTGTGGCGGCCGAGCGAGAAAATCGTGACGGG  
CACGCGGCTCGCGCCCCGGCGCCGATGCGCACGTGCGCCCGACGTGACCC  
CCGACGCCGCGACACGCGGCTGGCGGTGGACGTGTTACCTCTCGCCG  
CCCAGCGCATCGAGGCCGGCAGCGAGCGGTTCTGTCGTCTGGCAGCGTCC  
GAGCTCCGAGTCCCCGTGGCAACGGGTGCGAGTGCTCGACTACAACGCTG  
CCAGCCGAAGAGGCAAGCTGGCCGAGACGACCGTGCCGCATGCCAACTTC  
GAGCTGCTCATCACCGTCGAGAAGCAGAGCAGCCCTCAGTCTCCATCTTC  
TGCCGCCGTATCGGGCCGACGTCCGTGCGGTAACATCGCGCTATCAGCA  
GCGCTGAGCCCGCCAGCAGGCCCCAGAGCCCTGCCTCGATCGCCTTCTCC  
ATCATATCATCCCTGCGTACTCTCCAGCGACGGCCGCGTGAAGCAACC  
GCCGTGCGCGCGCGGCTCTACGTGCGCGACAGGAGAGCGTCTTGGCGCGG  
CCTGCGCATCGCTGGAAGGATCGGCGGAGCATGGAGAAAGAATCGAGGAT

5

178

10

5

15

20

10

25

15

30

35

40

25

45

50

55

CGCGATCTACGGCGCCATCGCAGCCAACGTGGCGATCGCGGCGGTCAAGT  
TCATCGCCGCCGCGGTGACCGGCAGCTCGGCGATGCTCTCGAGGGCGTG  
CACTCCCTCGTCGATACTGCAGACGGGCTCCTCCTCTGCTCGGCAAGCA  
CCGGAGCGCACGCCCGCCCGACGCCGAGCATCCGTTCCGCCACGGCAAGG  
AGCTCTATTTCTGGACGCTGATCGTCCGCATCATGATCTTCGCCGCGGGC  
GGCGGCGTCTCGATCTACGAAGGGATCTTGACCTCTTGACCCGCGCCA  
GATCGAGGATCCGACGTGGAACCTACGTCCTCTCGGCGCAGCGCCGTCT  
TCGAGGGGACGTCGCTCATCATCTCGATCCACGAGTTCAAGAAGAAGGAC  
GGACAGGGCTACCTCGCGGCGATGCGGTCCAGCAAGGACCCGACGACGTT  
CAGCATCGTCCTGGAGGACTCCGCGGCGCTCGCCGGGCTCACCATCGCCT  
TCCTCGGCGTCTGGCTCGGGCACCGCCTGGGAAACCCCTACCTCGACGGC  
GCGGCGTCGATCGGCATCGGCCTCGTGCTCGCCGCGGTTCGGTCTTCCT  
CGCCAGCCAGAGCCGTGGGCTCCTCGTGGGGAGAGCGCGGACAGGGAGC  
TCCTCGCCGCGATCCGCGCGCTCGCCAGCGCAGATCCTGGCGTGTGCGCG  
GTGGGGCGGCCCTGACGATGCACTTCGGTCCGCACGAAGTCCTGGTCGT  
GCTGCGCATCGAGTTCGACGCCGCGCTCACGGCGTCCGGGGTTCGGGAGG  
CGAGGGAGCGCATCGAGACCCGGATACGGAGCGAGCGACCCGACGTGAAG  
CACATCTACGTCGAGGCCAGGTGCTCCACCAGCGCGGAGGGCGTGACG  
CGCCGTGGAGAGACCGCGCGCGGCTCCGCCATCCTCCGCGGCGCCCGG  
CTCAGGTGGCCCTCGCAGCAGGGCGCGCCTGGCGGGCAAACCGTGCAGAC  
GTCGTCTTCGACGCGAGGTACGCTGGTTGCAAGTCGTACGCCGTATCG  
CGAGGTCCGGCAGCGCCGGAGCCCGGGCGGGCCGGGCGCACGAAGGCGCG  
GCCAGCGCAGGCTTCGAGGGGGCGACGTCATGAGGAAGGCCAGGGCGCA  
TGGGGCGATGCTCGGCGGGCGAGATGACGGTGGCGTTCGCGGCTCCCCG  
GCGCCGGCGCGCTTCGCGCCGCGCTCCAGCGCGGTGCTCGCGCGATCTC  
GCCCCGCGCCGGCTCATCGCTCCGTGTCCCTCGCCGGCGGCGCCAGCAT  
GGCGGTCTCTCGCTGTTCCAGCTCGGGATCATCGAGCGCCTGCCGATC  
CTCCGCTTCAGGGTTTCGATTTCG3CCAAGGTGACGAGCTCCGATATCGCG  
TTCGGGCTCACGATGCCGGACGCGCGCTCGCGCTCACCAGCTTCGCGTC  
CAACCTCGCGCTGGCTGGCTGGGGAGGCGCCGAGCGCGCCAGGAACACCC

179

55

5

180

10

15

20

25

30

35

GCGCACCCGGCGAGCCTCGGCGCGACCGGCGTCCACGAGCAGGCCTCGAT  
CCTCTCGCTGTACGACCCGTACCGCGCGCGCGCGCCGACGCGCGGCGGCC  
AGGTGCGCTCGTGGGAGGCGCTCTCCGCGCGCTTCGGCGGCGACCGCGAG  
GACGGCGGCGCTGGCCTCCGCTTCGTCTCCAGCCCACGAGCTCGCCCCCT  
5 CATGCGCGCGCTGATCGAGCGGTCCGGCGCAGGTTCCCCGGCGCGCGGT  
TCACCTTCTGGTCGCCGGTCCACGCCGAGCACGCGCTCGAAGGCGCGCGG  
GCGGCGCTCGGCCTCAGGCTCTTGCTCAGCTCGACGTCGACCAGGCCGA  
GGTGATCCTCTCGCTGGACGCGGACTTCCTCGCGGACATGCCGTTACGCG  
TGCGCTATGCGCGCGACTTCGCCGCGCGCCGCCGCCCGCGAGCCCGGCG  
10 GCGGCCATGAGCCGCCTCTACGTGCGGGAGGCGATGTTACGCCCCACGGG  
GACGCTCGCCGACCACCGCTCCGCGTGCGGCCCCCGGAGGTGCGCGCGG  
TCGCGGCCGGCGTCGCGGTGGGAACCTCGTGACGAGTCTTTGTCTTGCGC  
CCTGTCCGGGAATAACGGACACCTTATCGCGGGTCGCTCTTTGTGCGCGG  
CTTCTGTACCTCTCAGGACAGGTAGAAGAGGGAAGTACAGGGGCCCTTATGT  
15 TAACTGGGGATGCCTTCGGGACGGCCGCAAATATATCCTATCACCTCACT  
GGGTGTGGGGGAGCACCGCGAGGATGTACAACCTCTGTAACCTATGTGA  
GATAATGTGTGCAGTGATCTGAGACTTATTTGTGTGACCGAGACGTCTCT  
CTTATTGGTACGCATAGTATAATATAACACGTCTCATACATACTCCCGAC  
ATATCCGCGGTATGCGCGCACATAGAATAGGTGATGATAAATCCCTAGTG  
20 TGTGGAAGTAGAAGATGCGGGAGTTACCTGATATTTACGGAAAAAGTATT  
ATCTCAACTACCTCTCTGTTGAGACTATCACTTCGGTGTCGTTGTGCTGC  
TGGT,

40

or its complementary strand,

25

45

(b) DNA-sequences which hybridise under stringent conditions  
to regions of DNA-sequences according to (a) encoding proteins  
or to fragments of said DNA-sequences,

50

55

5

10

(c) DNA-sequences which hybridise to the DNA-sequences according to (a) and (b) because of a degeneration of the genetic code,

15

(d) allele variations and mutants resulting by substitution, insertion or deletion of nucleotides or inversion of nucleotide segments of DNA-sequences according to (a) to (c), wherein the variations and mutants offer isofunctional expression products.

20

11. Peptide encoded by a DNA sequence according to claim 10 selected from the group consisting of

25

Seq ID No 83

>Contig56\_003 2890 amino acids MW=307428 D pI=5.76 numambig=13

15 IRPRAAAVPMRSTVTGGVIAGPELGASYWADNLRQPVRFAAAAQALLEGGPALFIEMSPH  
PILVPPLDEIQTAAEQGGAAVGSLRRGQDERATLLEALGTLWASGYPVSWARLFPAGGRR  
30 VPLPTYPWQHERCWIIEVEPDARRLAAADPTKDFYRTDWPEVPRAAPKSETAHGSWLLLA  
DRGGVGEAVALSTRGLSCTVLHASADASTVAEQVSEASRRNDWQGVLYLWGLDAVVD  
AGASADEVSEATRRATAPVLGLVRFLSAAPHPFRFWVTRGACTVGGEPEASLCQAALWG  
20 LARVAALEHPAAWGGGLVDLDPQKSPTEIEPLVAELLSFDAEDQLAFRSGRRHAARLVAAP  
35 PEGDVAPISLSAEGSYLVTGGLGGLLVARWLVERGARHLVLSRHGLPERQASGGEQP  
PEARARIAAVEGLEAQGARVTVAAVDVAEADPMTALLAIEPPLRGVVHAAGVFPVRHLA  
ETDEALLESVLRPKVAGSWLLHRLRLDRPLDLFVLFSSGAAVWGGKGQAYAAANAFDLC  
40 LAHRRRAHSLPALSLAWGLWAEGGMVDAAHARLSDIGVLPMTATGPALESLERLVNTSAV  
25 QRSVTRMDWARFAPVYAARGRRNLLSALVAEDERAASPPVPTANRIWRGLSVAESRSALY  
ELVRGIVARVLGFSDFGALDVGRGFAEQGLDSLMALEIRNRLQRELGERLSATLAFDHPT  
VERLVAHLLTDVLKLEDRSDTRHIRSVAADDDIAIVGAACRFPGGDEGLEYWRHLAEGM  
45 VVSTEVPADRWRAADWDYDPDPEVFGRTYVAKGAFLRDVRSLDAAFFAISPREAMSLDPQQ  
RLLEVSWEAIERAGQDPMALRESATGVFVGMIGSEHAERVQGLDDDAALLYGTTGNLLS  
30 VAAGRLSFFLGLHGPTMTVDTACSSSLVALHLACQSLRLGECQALAGGSSVLLSPRSFV

50

55

5

182

10

AASRMRLSPDGRCKTFSAAAADGFARAEGCAVVVLKRLRDAQRRDPILAVVRSTAINHD  
GPSSGLTVPSPGAQQALLRQALAQAGVAPAEVDFVECHGTGTALGDPPIEVQALGAVYGRG  
RPAERPLWLGAVKANLGHLEAAAGLAGVLKVLLEHEQIPAPPELDELNPHIPWAELPV  
AVVRRAPVPWRGARPRRAGVSFGLSGTNAHVVLEEAPAVEPVAAAAPERAAELFVLSAKS  
5 AAALDAQAARLRDHLEKHVELGLGDVAFSLATTRSAMEHRLAVAASSREALRGALSAAAQ  
15 GHTPPGAVRGRASGGSAPKVVFVFPQGGSQWVGMGRKLMAEPPVFRAALEGCDRAIEAEA  
GWSLLGELSADAAASQLGRIDVVQPVLFAMEVALSALWRSWGVEPEAVVGHSMEVAAA  
VAGALSLEDAVAIICRRSRLRLRISGQGEMALVELSLEEAEALRGHEGRLSVAVSNSPR  
STVLAGEPAALSEVLAALTAKGVFWRVKVDVASHSPQVDPLREELIAALGAIRPRAAAV  
20 PMRSTVTGGVIAGPELGASYWADNLRQPVRFAAAAQALLEGGPALFIEMSPHPILVPPLD  
10 EIQTAAEQGGAAVGSLRRGQDERATLLEALGTLWASGYPSWARLFPAGGRRVPLPTYPW  
QHERYWIEDSVHGSKPSLRLRQLRNGATDHPLLGAPLLVSARPGAHLEQALSDELSYL  
25 SEHRVHGEAVLPSAAYVEMALAAGVDLYGTATLVLEQLALERALAVPSEGGRIVQValse  
EGPGRASFQVSSREEAGRSWVRHATGHVCSGQSSAVGALKEAPWEIQRRCPVLSSEALY  
15 PLLNEHALDYGPCFQGEQVWLGTGEVLGRVRLPGDMASSSGAYRIHPALLDACFQVLTA  
LLTTPESIEIRRLTDLHEPDLPRSRAPVNQAVSDTWLWDAALDGGRRQSASVPVDLVLG  
30 SFHAKWEVMERLAQAYIIGTLRIWNVFCAGGERHTIDELLVRLQISVVYRKVIKRMEHL  
VAIGILVGDGEHFVSSQPLPEPDLA AVLBEAGRVFADLPVLFEWCKFAGERLADVLTKCT  
LALAILFPGGSFDMARIYRDSPIARYSNGIVRGVVEAARVAPSGMFSILEIGACTGA  
20 TTAAVLPVLLPDRTEYHFTDVSPIFLARAEQFRDYPFLKYGILDVDQEPAGQGYAHQRF  
35 DVIVAANVIHATRDIRATAKRLLSLLAPGGLLVLVEGTGHPiWFDITTGLIEGWQKYEDD  
LRIDHPLLPARTWCDVLRVGFADAVSLPGDGSPAGILGQHVLRSRAPGIAGAACDSSGE  
SATESPAARAVRQEWADGSADVHRMALERMYFHRRPGRQVWVHGRLRTGGGAFTKALAG  
40 DLLLFEEDTGQVVAEVQGLRLPQLEASAFAPRDPREEWLYALEWQRKDPIPEAPAAAASSSS  
25 AGAWLVLMDDGGTGAALVSLLEGRGEACVRVIAGTAYACLAPGLYQVDPAPQPDGFHTLLR  
DAFGEDRICRAVVMWSLDATEAGERATAESLQADQLLGSLSALSIVQALVRRRWRNMPR  
45 LWLLTRAVHAVGAEDAAASVAQAPVWGLGRTLALHPELRLCTLVDPNPAPSPEDAAALAV  
ELGASDREDQVALRSDGRYVARLVRSSFSGKPATDCGIRADGSYVITDGMGRVGLSVAQW  
MVMQGARHVVLVDRGGASEASRDALRSMAEAGAEVQIVEADVARRDDVARLLSKI EPSMP  
30 PLRGIVYVDGTFQGDSSMLELDAERFKEWMPKVLGAWNHLALTRDRSLDFFVLYSSGTS

50

55

5

183

10

LLGLPGQGSRAAGDAFLDAIAHHRCKVGLTAMSNWGLLSEASSPATPNDGGARLEYRGM  
EGLTLEQGAALGRLLARPRAQVGVMLNLRQWLEXXXXXXXXXXXXXWYNLLII IQYTK  
VPPQGPFRL\*

5 Seq ID No 84

15

>Contig56\_027 700 amino acids MW=80569 D pI=7.02 numambig=0  
MNMELNITSKSNPFGDTTAENDKKMLSNAFIETADFRTLIETDDRTIVVGRRTGKSALF  
IQLNEHWKKDKKILILSFSPPDSQIIGFRSMLKPFTGSFNLARAATRLWRYAMLMEIAS  
YISSHYKLSSQISSETLLNEHLKKWNSAQGDILRKCLVAKEYLDENNPEESIGDLQFNL  
10 NISEIENNIVSLLERSDRKVILMDKLDEAYEPDNIGIGIIAGLAYASIELNQAKACIRP  
IIFLRDNIFRSLSKEDPDYSRNIQGVIRLHWDWAQLMLLSAKRMKVAFKLDIEKDQVRW  
DRCTADDLKGNGFKRCLQFTLYRPRDLLSLLNEAFFSAFRENRETIINTDLEYAAKSIS  
25 MARLEDLWKEYQKIFPSIQVITSAFRSIEPELTVYTCLKKIEASFELIEENGDPKITSEI  
QLLKASGILQSLYSVGFVGIRDKNTSSYSFCHDGRTPDKGFESNEKLLIHPCYWLGLNLN  
15 RNALAPEEAEEINDEYDINIISDNSAIRNKTIGQITTHLDQIPIGNEGATEFEQWCLDAL  
RIVFASHLTDIKSHPNGNAVQRRDIIGTNGGKSDFWKRVL EDYKTRQVVFDKKNFEEELGP  
30 SEYRQLQSYLTGPYKGLGFIINRDESEVLKSGKDL DWTKEYQSHNSLI IKLPAKYISKL  
LQKL RNPEKHDAIDRQMGKLLTLYETSYMAIKSTQKKRRK\*

25

30

20 Seq ID No 85

35

>Contig57\_001 372 amino acids MW=38411 D pI=12.39 numambig=10  
MLTSXXXXXXXXXXLLAYRCATAARGAGRIRHHRQYAQRRRTTVVLYARAAGTLPDRRHG  
LLVVARGHPPCLPQPARSRERSRAGGRRQHAPFVQDDDNAGAHGAVARWPLPDIRRLGQ  
40 RVRPWGGLRYGRAQTALRRPATRRSDLGSDPGFGHESGWFPVDRVDGTQCARSGGALARGA  
25 AERSRRRRGHRLCRDPRNGDLARRPDRGRGAACRVGAGAGRWEPLRAGRSEDKPRPPGGR  
CRRCGFDQGGAGSAPRTDPAKPPFPHAQSADPDRGDRARAGDGAGAVAAGGPTALRGGER  
VRFQRHQRPCRAGGGAGHGARTGDAGALSGAFGAVGEERRRAGRTGGAALSAHRRVPGAG  
45 SRRRRVQPGIDA\*

50

55



5

184

Seq ID No 85

10

&gt;Contig57\_002 2259 amino acids MW=238258 D pI=5.92 numambig=0

15

5

MSYTLGLQGPCLTVDTVCSSSLVAIHLACRSLRARESDLALAGGVNMLLSSKTMIMLGRI  
QALSPDGHCRTFDASANGFVRGEGCGMVVLKRLSDAQRHGDRIWALIRGSAMNQDGRSTG

20

10

LMAPNVLAQEALLREALQSARVDAGAIGYVETHGTGTSLGDPiEVEALRAVLGPARADGS  
RCVLGAVKTNLGHLEGAAGVAGLIKAALALHHELI PRNLHFHTLNPRIRIEGTALALATE  
PVPWPRAGRPRFAGVSFGLSGTINVHVLEEAPATVLAPATPGRSAELLVLSAKSAAALD  
AQAARLSAHIAAYPEQGLGDVAFSLVSTRSPMEHRLAVAATSREALRSALVAAQGGTPA  
GAARGRAASSPGKLAFLFAGQGAQVPGMGRGLWEAWPAFRETDFRCVTLFDRELHQPLCE

25

15

VMWAEPCSSRSSLLDQTAFTQPALFALEYALAALFRSWGVEPELVAGHSLGELVAACVAG  
VFSLEDAVRLVVARGRLMQALPAGGAMVSTAAPEADVAAAVAPHAALVSTAAVNGPEQVV  
IAGAELKFVQIIAAFAARGARTKPLHVSHAFHSPMLDPMLEAFRRVTESVTYRRPSIALV  
SNLSGKPCCTDEVSAAGYVWRHAREAVRFADGVKALHAAGAGLFVEVGPKPTLLGLVPACL

30

PDARPVLLPASRAGRDEAASALEALGGFWVVGSVTWSGVFPSSGRRVPLPTYPWQRRERY  
WTEAPVDREADGTGRARAGGHPLLGEVFSVSTHAGLRLWETTLDRKRLPWLGEHRAQGEV  
VFPGAGYLEMALSSGAEILGDGPIQVTDVVLITETLTFAGDTAVPVQVVTTEERPGRLRFQ

35

20

VASREPGERRAPFRIHARGVLRRI GRVETPARSNLAALRARLHAAPAAA IYGALAEMCL  
QYGPALRGLAELWRGEGEALGRVRLPEAAGSATAYQLHPVLLDACVQMIVGAFADRDEAT  
PWAPVEVGSVRLFQSRSPGELWCHARVVS DGQQAASSRWSADFELMDGTGAVVAEISRLVVE

40

25

RLASGVRRRDADDWFLELDWEPAALGGPKITAGRWLLLGEGGGLGRSLCSALKAAAGHVVV  
HAAGDDTSTAGMRALLANAFDQGAAPTAVVHLSSLDGGGQLGPGLGAQGALDAPRSPDVDA  
DALESALMRGCDVSLVQALVGMDLRNAPRLWLLTRGAQAAAAGDVSVVQAPLLGLGRT  
IALEHAELRCISVDLDPAEPEGEADALLAELLADDAEEVALRGGDRLVARLVHRLPDAQ

45

RREKVEPAGDRPFRLEIDEPGALDQLVLRATGRRAPGPGEVEISVEAAGLDSIDIQLALG  
VAPNDLPGEIEIPLVLGSECAGRI VAVGEGVNGLVVGQPVIALAAGVFATHVTTSATLVL  
PRPLGLSATEAAAMPLAYLTAWYALDKVAHLQAGERVLIHAEAGGVGLCAVRWAQRVGAE  
VYATADTPENRAYLES LGVRYVSDSRSGRFVTDVHAWTDGEGVDVVLDSLSGERIDKSLM

50

VLRACGRLVKLGRRDDCADTQGPLPPLLRNFSFSQVDLRGMMLDQPARIRALLDELFGLV  
AAGAISPLGSGLRVGGSLTPPPVETFPISRAAEAFRRMAQQQHLGKLVLTLDDEVRIRA  
PAESSVAVRADGTYLVTGGLGGLGLRVAGWLAERGAGQLVLVGRSGAASAEQRAA VALE

55

5

185

10

AHGARVTVAKADVADRSQIERVLREVTASGMPLRGVVHAAGLVDDGLLMQQTTPARFRTVM  
GPKVQCALHLHTLTREAPLSFFVLYASAAGLFGSPGQGNAAAFLDALSHHRAQGLP  
ALSIDWGMFTEVGMAVAQENRGARQISRGMRGITPDEGLSALARLLEGDRVQTVIPITP  
RQWVEFYPTATAASRRLSRLVTTQRAVADRTAGDRDLLEQLASAEPSARAGLLQDVVRVQV  
5 SHVLRLPEDKIEVDAPLSSMGMDSLMSLELRNRIEAAALGVAAPAALGWYPTVAAITRWL  
15 LDDALVVRLLGGGSDTDESTASAGSFVHVLFRFPVVKPRARLPCFHGSGGSPGFRSWSEK  
SEWSDLEIVAMWHDRLASEDAPGKKYVQEAASLIQHYADAPFALVGFSLGVRFVMGTAV  
ELASRSGAPAPLAVFTLGGSLISSSEITPEMETDI IAKLFFRNAAGFVRSTQQVQADARA  
DKVITDTMVAPAPGDSKEPPVKIAVPIVAIAGSDDVIVPPSDVQDLQSRITTERFYMHLLP  
20 GDHEFLVDRGREIMHIVDShLNPLLAARTTSSGPAFEAK\*

20

10

Seq ID No 87

25

>Contig57\_027 419 amino acids MW=46737 D pI=5.09 numambig=0  
MTIQEQANQSETKPAFDKPFAPGYAEDPFPALERLREATPIFYWDEGRSWVLTRYHDVSA  
15 VFRDERFAVSREEWESSAEYSSAIPELSDMKKYGLFGLPPEDHARVRKLVNPSFTSRAID  
LLRAEIQRTVDQLLDARSGQEEFDVVRDYAEGIPMRAISALLKVPAECDEKFRFGSATA  
30 RALGVGLVPQVDEETKTLVASVTEGLALLHDVLDERRRNPLENDVLTMLLQAEADGSRLS  
TKELVALVGAI IAAAGTDTTIYLI AFAVLNLLRSPEALELVKAEPGLMRNALDEVLRFDNI  
LRIGTVRFARQDLEYCGASIKKGEMVFLIPALRDGTVF SRPDVFDVRRDTGASLAYGR  
20 GPHVCPGVSLARLEAEIAGVTIFRRFPKMKLKETPVFGYHPAFRNIESLNVILKPSKAG\*

35

Seq ID No 88

40

>Contig57\_043 492 amino acids MW=52617 D pI=11.54 numambig=0  
MAARARKSCRAGSRPAPMRTSPPTSTPTPRPRGWRWTSFTSRRPSASRPAASGSSSGSV  
25 RAPS PRNGSECSTTTLPAEEASWPRRPCRMP TSSCSSPSRSRAALSLHLLPSSGRRPS  
GNIALSAALSPPAGPRALPRSPSPSYHPCVLLQRRPRRSNRRAGAALRARQESVLARPAH  
RWKDRRSMEKESRIAIYGAIAANVAIAAVKFIAAAVTGSSAMLSEGVHSLVDTADGLLLL  
45 LGKHRSARPPDAEHPFGHGKELYFWTLIVAIMIFAAGGGVSIYEGILHLLHPRQIEDPTW  
NYVVLGAAAVFEGTSLIISIHEFKKKDGGYLAAMRSSKDPTTFTIVLEDSAAAGLTIA  
30 FLGVWLGHRLGNPYLDGAASIGIGLVLA AVAVFLASQSRGLLVGESADRELLAAIRALAS

50

55

5

1.86

10

ADPGVSAVGRPLTMHFGPHEVLVVLRIEFDAALTASGVAEERERIETRIRSERPDVKHIY  
VEARSLHQRARA\*

12. DNA sequence according to any of claims 1 to 5 wherein the  
DNA is selected from the group consisting of

15

(a) the following DNA sequences:

Seq ID No 89 (>Contig10)

20

10 GGTAGTGAAATATGCTGTATTCAACAGAAAGCTTGATGAATTGATCTAGA  
AAGTAGAGCGAGAGAATCAAGTAAGATAGTAGGATGCATTATAAATATAG  
AATATATACTGCATACGATGACAGCATGCGCACGAATAGAATGCATAAGA  
25 GGCAAGCCAATAACCAAAAGTGGAGCCAGAGGAGATAGTCTCGCCAGTAG  
AAATAATGCTCAGCCAAGCGAGGTTGGACATATCAGTTCCAGAGTAGGTC  
15 TCAACCCCGTATATGAGTCCAATGAAGCCTGTCTCATCCAGTTAACGGCC  
TTTGTAGCAGAGAATCCTCCCTATTTTCGGAGAGGACGCGTCGAATATAA  
30 AGCAGGTCCAAAGAAGCAAGCAATAGCCAAAAGTTTGAAAGGTTAGTACG  
AGCAGCGGCTGGAGGACACTATGGTCGTGCAACGGGGGTAAAGGTTTCA  
CGTATTGTAGCAGAGCACGTCAGAGGGTTATTCGTGACATTCGAGGCCAA  
20 CGAGGCGGTAGGACTTCGTAAGCGCATGACCATCCCGGTCACAAACGTAG  
35 TCGGAGCGCCTCGTCAAGCTCAACAAGGCCCTAGAACGCGCGGCGCAGA  
TCGACCCTTTTAAACGCCGGCACCAGCCGACCGTCCTGCCAGGTTGT  
AAAGCGCTCCATCGGCCGACTTATGGCACTCGAGCCAAATCGCCCGGTTTC  
40 CCCATCGGTGAGCGCAAACGGCCCCCGGGCGTCGCCACCCGCGCGGAC  
25 GAGGGGCGGTCCAGACGGGTGATCTCTCTCGTGAGCTCGCGGAGAGAGCC  
TCCTCGCAAGATCGATGTCAGCGGGATCGCGCGCCCGTCCGCACCTGAA  
ACCGGTGCTGGAGCTCGACGGCAGCGAGGGAGTCGAGGCCGAACCGCGAT  
45 ATCGGCAGCGCGTCGTGATCTGCCCCGGCGTCCAGACGAAGCGCGCGGGC  
GAGGGTTCAGCGCAGCGGTCCAGCAGGCTCCGGCCGAGGGCTCTTCGG  
30 TCTCCGGGGGCGCGTCGTCCGGGGCGAGGCGTCGTGAGGAGCTCCGGC

50

55

5

187

10

5

15

20

10

25

15

30

35

40

25

45

30

50

55

GCGAACGCGACGTGGCGCTCGCCGAGCGCGTCCTCGAGAAAGGCGCGCCG  
GCACTCCCTCCGGCGGACCTTCCCGCTCGACGTCTTCGGCAGCGCGCCCG  
GCGCGATCAGCGCGACGGCGTGCGCGACGAGCTGGTGCTCGGCGGTACAC  
GCCTCGCGCACGGCCGCCACGATCTCGCGCGGATCCGCGGCCACGCGCGG  
GTGACCTCGCACACCACGGCGAGGCGCTCCTCGCCCTCGTGCTCCACGG  
AGAACGCGGCGCTGCAGCCCGGCCGACGGCGCGATGGCTGCTCTCGACG  
GTCTTCTCGATGTCTCGGGGAAGTGGTTGCGGCCTCGAAGGATGATGAG  
GTCTTTCGACCTCCCCACCACGAACAGCTCGCCGCCCCGGAGGAAGCCGA  
GATCTCCCGTGCGCAGGTAGCGCGGCGCCGCGCTGCCAGCGAGCGTGCC  
CCGAACGTGGCCTCCGTCTCCTCCGGGCGCCCCAGTAGCCGACGGCTAC  
GCTGGGCGCGGACACCCAGATCTCCCGATCTCCCCCGGCCGAGCTCGT  
TCCCCGCGGGATCGACGATCGCGACCGCCCGGGATCGAGCGCCCGACCG  
CTGCCGACGAACACGCGCGCGCCCTCCGCGCCGACGCGACGGCGCGCCC  
GAGTCCACCTCCTCGGGGCGAGGCGCGCCAGCACCGGCGCCTCGGCCC  
GCGCTCCGCGCGCTCACGATGAGCGTGGCCTCGGCGAGCCCGTAGCAGGGA  
TAGAACGCCTCTCGCCGGAACCCGCTGACCGCGAAGGCGCGCGGAAGCG  
ATCGAGCGTGTGCGGCGCGCACCGGCTC3GCGCCCGTGAACGCGACCTCCC  
ACGACCGCAGATCGAGCGCCGCTCGCTCCTCCTCCGAGCTCTTCCGGACG  
CACAGGTGCTATGCGAAGTTGCGGCGCGCGCTCACCGAGGCGCCGAGCGC  
CGACACGGCGCGGAGCCACCGCATCGGCCTCTGCAGGAACGAGAGCGGCG  
ACATGAGCGCGACCGCGATCCGCCGCTAGAGCGCCTGCAAGATCCCGCCG  
ATGAGCCCCATGTCTGTGATACGGCGGCAGCCAGATCACCCGACCGGATC  
CGGGCTCGTCAGGTGCAATCCATGCGCGATGAGCCGCGAGTTGTGCAGCA  
GATTCCTCGTGGGTGAGCATCACCCCTTGGGCTCGCCGGTCGAGCCGGAG  
GTGTATTGAAGGAACGCGACCGACTCCGGCCGAGCGCCGCGCCCGGCC  
CTCGATCGGGCCCGGCGACGGCCGTCGGTCGCGATCCACCGAGCCGCT  
GCAGCGCGGCGCCGCGCGCTGGCCGCGAGGACGCCACGATGCCGGCG  
ACGGCCGATGACGTGAGCGCCGCTCGGCGCGCGCTCCGCGACGATGGA  
AGCGACGCGCGGCAGCGTCCGCTCGAGCCGGCCGAGATCCGGCGGATAGG  
CGGGCACGGTCCGGACTCCAGCGTAAAGACACCCGAAGAACGCGGTGATG

5

188

10

15

20

25

30

35

40

45

50

55

TACTCGATCCCCGGCGGATACAGCAGCAGCGCGGGCCCCGGGGGCGAC  
GCCCCGATGCCTGCAAGAGGGCCGCGACGGTTCGCGCGCGCTCGTCAATTT  
CCCCGAGGGTCACCCAGGTCGCCCCGGCCTCGACGTCGCCGGA CTCAAGA  
AAGCAATAGATTGGCGGGCGGGCTCAGCTTCGGCCCCGCTGGCGCAAGAG  
5 GTCGATAACGGTGAAGGGCGGTTCCGTTCCGTTCCAATGCAAGAA  
AAGCATCATTGTAACAGACCCCTCCGCCGCGGAGATAGCAGCTTGTC  
CGCTGCGACACAACCGCCGCGGACGCGCGTGGCACGGCGGGATCCGGGC  
GTACTCCACCTGCACTTCCCGTCGCGTCACGCTCGCTCCGCCGCGGGTG  
TCGTGAACCACCGCCACAGCGACACGCGCCCTGGCTCCGAGGGCGGCGT  
10 CCACGTGGTCCCGTTGCCCCCGGGAGGCGCCGGTCGTATCGCTACCA  
GGCGGCGCGCCCCGTGAACTCGCCACCGTCCGTGTAATAGTCGACCCAG  
ATCGCCTCGCGCGCCGGTGGACCGCGAGCCCGGCGGCTTCCTCGTCCAC  
CTCGGCGGCTTCTCGGGCACGACAGCCTCAATCTCATAGGTCGTGCACT  
25 CGTCTCGGCCGCTCGGTCCGGCCGCGAGCCTTGGGCTGCTCCTCGGAC  
CGAACGCACCGCTTCACGACGGGCAAACCGTCCTCGCCCCGGCGCGACCTC  
ATTGCCATCGAGCTTCAGCGTGAAGCCGTCGATGGGCGGGTTCGTGTTCA  
30 GCCGCTCCTTCTTGAAGACATAGACCTGCGTGTAGCCACGACGAAGCTG  
TCCGGACCGAGCACCGTCCCGTCGTCGCGACGCACTCCAGCGGAAACCC  
GGCCGTTTCGGGCGCCGAAGCCACGCGTGTGTCGCGCGCACACGGCGA  
20 ACAGCACGTAAGCCGACGAGTACACCGTCCCCGTCTCGGTGGGCCTCGCG  
TCCTTGAGGATCTCCTTGGGCAGCTTCACCCGAACGAGACCGCATCGGG  
CTCGCCGCTCTTCTCCGGACCGATCTCCTGCTGCGCGAAGGGGACGGTGC  
GCTCCCCGTGCGCGCTGCCGCTGCCGCCGTGCGCGCTGCTGCCGCCGTG  
40 CCGTGGCACCGCCACCGCCCGCCCGCCAGCGCCACCATTGCCGCCTTC  
25 GCCGCTCCAGCGCCACCACTGCCGCCGTGCGCGCGCTGCCACCATCGC  
CACCCTGCCGCCCGCCGCTTCCGCCGTGCCGGCCGGCACCGCCTCC  
CGGATTCGCGACGATTCCCACCGCATAGGTGCCAGCCACTGCGGGATGC  
45 ACCCGAGGTGCTCGTCCACCCCGACCGGCGGATTACGCAGCCGCCACC  
CACGTGACCTCGACCTTCCGCGGCGCGCCGCTCCGCGCCTTTCGCGTC  
30 GGCGTACGTCATCCGGAACGTCACGAGCTCTCCGCCGCCGCTACGGCT

5

10

15

20

25

30

35

40

45

50

55

TGTCCGCCCTCACGGCGAGGACGCGGAGCCCCCTTACCTCGGACGAAGGG  
GCCATGTCGCTCCCGGCGCAGCAAGGGATGCCCACGGCCAGGGTCGAGAG  
CGCGGCCAGCAGCGCGCGACGAGCGGGCAGTGCGGTCCGTTTCATCAGAA  
ATCTCCTCGCAGCCCGAGCGTGGGCAGGAAGGGGAGCCCCGTCACGTACT  
5 CCCGCTTCGTGTAGTTGAAGTTGTAGCTGATGCCCTCCGCGAGCCATGTAA  
TTGTAGACGTTCTGGATATCGAGGTAGAGCCCAGCTGCCACCTCTTGAA  
TTTCCACGTCTTGTGCGCGCGGATGTCGAGCTGGTGAACAGCGGCATCC  
GCTCGCTGTAGTCACCCCCGAGCGGGATCGGCGAATACCTCGCCGAGGAC  
GCGTGGTAGATCGCGTTCACCCGGTTCGGATTGCACCCCTTCTCCTCCGG  
10 ATCGCAGACATAGGGCGTCTGCAGGTTGCCCGACACGAGCCGGAAGCGCG  
CGCCAGCTCCCAGCCCCGGCCGAGCCGAGGCTCCCGAGCACCCGTCAGC  
ACGTGCGTCTGATCGAACTGGGTGAGGTGCTCCTCCTCGTCGGGGCCGTC  
CTTGCGCACCGACCGCGAGAGGGTGTACGCCGCCAGCCGAAGAAGCGCT  
25 CGTCCGGCTTGTACTTCAACAAGAGCTCGCCGCCGACCGCGTATCCGGTG  
CCATCGTTGGCATAGTCGTCTTCTCCGGCGAGAAGACGACCAGCCGATC  
GAGCTGCTTGTAGAACCCGTCCAGCGTCACCTCGATCTGCGGCGTGATCT  
30 CCTGCTCCACGCCGAGGCCGTAATGCACGGCGCGGTTGACTTGAGCTCC  
GCATTGCCGAACGGCTCGATGCTCTCCGCGAACTGCGGCGCCTGATAATA  
AAGGCCACGCCCCCTTGGCCGTCTCCGCGGAAGCCGCTCCGATGT  
20 CGTAGCCGCGGTTGACCCGCGGGCTCACGTGAGCGTCTGCGTATCGAGC  
GCGTAGTCGACCCGCACCCGGGACGATCCGCGCCCGGGCGAGGGGAC  
GACCTCGAGCTCGGCATACGCCGCGGGCCGCGAGTACGCGCCGTCGAACG  
ACCGATCCTGGAACGGGTACGTCGAGAACGGCTGGTTCGACGGGTGGCCC  
40 GCGGGCTGCTGCGACGCGCGCGGATGTTGACCGTGGCGACGCCGCCGA  
25 GAGGTCGGTGCCGACGTTTCATCGTGAGGTACCGCGGAACCTGTGCGAGA  
GCTCCAGCCGCGAGGTCGAGCGAGGTCGAGACGACGTTGAAGGCGAGGGGA  
GAGATCTCGAAGTCGGCGATGTCCCGGCCGAGCGCCATCGACCACAGCAG  
45 CCGATCCCGGCTCCCGATCCGGTTCTCGTAGCTGAGCTGGAAGCGCTGGA  
AGGCGGTGTGCAGCCCGAAATCGCCCGTCAGCGCCGCTCGTCCTCCGGC  
30 GGCTTGTCCAGGGTGATCTTGAAGGCGTCGTCCGATCCGTAGAAGCTCGC

5

190

10

5

15

20

10

25

15

30

35

20

40

25

45

30

50

55

GCCACGCGCTCGCTCGCGGAGGGGCGGCCCTCGAGGACGAACTGGTAAT  
CATAGTAGACGGGCGCCTGCGTGACGCTGGAGCCCGCCTCCTTGAGCACG  
GGCCCGAGCCACGCGTCGACCCAGCTGCGGCGGCCCGCCGCGATGAACGT  
CCAGTCCTTGAGGAACGGGACGGGGCCCTCGAGGAGCACGCGCCCGTCGA  
TGAGGTCGAGCTGGACCACGCCGTGGTACTTGCCGTCCTGCTTCGGCGAG  
CGGAGCCCGACGTCGACGATGCCGCCCATGGCGCGGCCGTACACGGCGCT  
GAAGTTGCCCGGATAGAAGTCGATCTTCTCGAGCATCTCGGTCGGCACGA  
CCGAGGAGAGGCCGCCGAAGTGGTAGATGATCCGCACCGGGGTGCGATCG  
ACGAACGTGAGCGTGTCCTGGGGCGCGGACCCGCGCACGATGAGCAGCCC  
GAAGCCGCTGCGCGGACGCCCGGCAGGCTCTGCAGCGACCCAGCGCGT  
CGCCCGCGGTGCCGGGATGCGGTGATCTCGCGGCGCTCGATCGTCCTC  
CGCGTCACCTCGCGCGGCGGGCGCTCGCCCTGCACGGTCACCTCGATGCC  
CGGCGCCTTGCCGTCCTGCGGCGGGCGAGCGAGATGCGGTACCGCACCT  
CGATCGCCTCGCCGGCCGCGATCTCCTCCTCGGCGGCGAACGGCTCGAAC  
CCCGCGGCGGCGACCTCGACCGGTAATTGCCGGGGGGGAGATTCTTGAA  
GCGGAATTGCCGCCCTGGTCCGTCTTCGCCTCCTCGCGGCCCGCTCGG  
GGCGCACGAGGGTGACCGCATGTCCGGAGCGGCTCGCCGGTGCCCGCG  
GACAGGACGGTCCCGACCACCGTCTCGACGTGCGCGGGCGGCGCCGGCGC  
GGCCGCATCGGCGGGCTTGGGCGTGAGCGTGAACGCGTACCGGTAGAGGA  
TGCGCGCCCGCGGGCGTGCCGTCCGGGCGCCGCGCCGGCGCGAACTCC  
AGGCCGGGCGCGGCCCGGAGCGCCGCTCGTTGAAGCCGTGCCCGCCGG  
GCGTCGCGACCTCGGCCCTGGTGACGCGCCCGGTCTTGTCGATGTCGAGC  
TTGAGGATGACGCTGCCCTCGACGCCGGCGCGCTGGGCCTCGATCGGATA  
CGCGGGCGGGAGTACTTGATCAGCGTCGGCGGGCTGATGGCGGCGGGCG  
CCGGCGGGGCGCGCCCGGCTGAGGGACGACGACCGCGCCGGCGCGCCG  
CGGGGACCGAGGCGCCCGGAGTCGCCCTCGGCGGCGAGGAGGCTC  
GGGCGGGGCGCGCCGGCGGGCTGCGCGCGCGCTGCGCTGCCGGTCATCG  
CGACCGGAGCAGCAGCGCTTCCGAGACGACGAGCGCATCACGAGGAC  
GCTGTGGAAGGCATGCGGCCCGCCCTCTCGCATGGCGAGGCCGAGGCGGA  
AAGACGCATCGCGCAGCCAGGACCCTGCTTACATTGCTTCACACAACGG

5

10

15

20

25

30

35

40

45

50

55

GCGCCGCGCGCTCCCGGGCCGCGCAGCGCAGGCGGCGCGCGCCCCG  
CGGGCGGCGCGATCGCGAGCGGCGCGCGGTGCGATCAGCCGCCGACCTCG  
GCCACGAACCGGCTCACGTCGTCGCTGTCGCCCACGAGCAGCAGCGTGTC  
GCCGTCGCGGATCACGTAGTCCGGTGTGGGCGCCTCGAGCCGCGGCTTGT  
5 CGCCGGGCCGCTTGTTCGTGTGCGGCCGCACACCGAGCACGTTGATGCGG  
TACCGCTGGCGGATCTTCGAGCCGGCCAGCGTCTGCCCCACCAGCGGCCC  
GTGGGCGTTCCAGGGGACCACGCGGTAGTGCTCGCGAGGTCGAGGAGGT  
CCTGCGCGAGCGGCATGGTGATGTGCGCGCCGACGCGGCGGCCATCTCG  
GTCTCGAGCTGGATGACGCGGGTCGCGCCACCGCGCGCAGGATGTGCGC  
10 CTGGCGATCGGTGGCGGCGCGCGGATGATCTCGCGCACGCCCATCCGGA  
CGAGGGAGGCCACGCAGACGACGCGGCTCGAAGTGCTCGCCGAAGGTC  
ACGATCGCGGTCTCCACGTACTGCGCGCCGATCCCCCTCAGCACCTTGTG  
GACGGTGGCGTCGCCGACGAACGCGGCCGAGGTCTTGTCTTCACGGCGT  
25 CGACGGCCTCCGGTTGTTGTGACCGCGATCACCTCGGCCCGGTTCTTC  
15 CAGAGGGTCTCGACGACCGACGTGCCGAACCGCCCGAGCCCCGATGACGA  
GGACGCTCTTCGATTTCATGGTCTCCGGTCGCGCGCCGCCCTCGGGGC  
GCGGCGCCGGAAGGTCTCACGGATGCGCCGGGAGCGCCACGGTTCGGCG  
30 CTGCCGTCCTCGCGACGGCGCGGCCCGCGCGGGCCGCGCGCGCGCGCT  
CAGTAGAGCTCGTCTGGTGCTGCCACCGCTCCGAGATCCACGCCTTCAG  
20 GTACGCGATCTCCTCCTCGTACGTCTGTAAGTCGTGCGCCAGCTCCAGC  
CCTCGTAGCTCCGGTACGCCTCGCCCCACCGCGCCTCGTCCCGGCGCGCG  
35 CTCGCGTCGATGCGCTCCACGTAGCCGTCCACGATCGCGTGGATCTCGGC  
CTCGCGGAGCGCGCCGCGCAGGACCTGATCGTAGCGGGCGCGCAGCGGGT  
CGCCGATCGACGGCTCCTCGAGGAGGCGCTCGAAGAGGAGGTTACGTCG  
40 CGGTAGTCGACGCGATCCGACGCCGGCTCGCGCTCGGTCTCCACGACTG  
25 GCCGAAGCTCGCGTTGAAGTCCCACGGCGCGTAGCGGAATACGCCGTCCG  
CGGCCGGATCGCGGTAGTGGTAGCTGTTCTTTCCGGCCGAGTCGTTGGCC  
45 ACGATGAACGTGACGAAGATCCACCAGTCCTCGTAGTCGCGCAGATCGAT  
CCGCGACCCGATCTCGGCGGCGAACGTGGCGTCGTGCGACTCGGCCACGA  
30 AGCTCACGAGATCTTCCAGATCCGAGAACGCCTCCGGCTCGCCCTCGGCC



5

192

10

GGCGCCCTTCCTTCTCTCGAAGCCGTCGTGCAGCGTGTCTTGGGGTTC  
GCCGGACCGGTCCGTACGCGCAAGTTCGCGTCGTGGCTGACCGCCTTGT  
AGAGGTTGCCGTCTCGCGGTAGCCGTGGTCTCCATCAGGTAGCCGTCTG  
ACGTGATCCGCGACGGTGTAGAGCCCCGCTACTCCCCGTGAGGTACAG  
5 GACGGCGCTGTAGGTCTTGATCTGGATGTGCTCGGGATCGAGGCGGTTC  
AGAGGTCATAGGCGAGGCGCTGCCGGACATAGGAGTTGTCGTGCAACGTC  
15 GTGATGAGCACGACCTTGCGGCGATCGGTGAAGCCGCCCGCTCGTCGGG  
CTCGTTGAACTTGTCTCTTGGGGAACTTGAGGGTGTAGCTCCGCTTCG  
GGTACGAGAGCGAGCTCTCGCCGCGGAGCTCCGCCTCCGCGGCGTACGTG  
20 TGGCCGCGGTAGATCACCGTGGCCGGGGCGTACTCCTTGTCCTCGGGGAC  
10 GGGCGAGAGGAAGAGCACCGGCAGGCCGTACTCCTCGGGGTAGCGGGTCG  
GATCGACGACGGGCACGTTTCGACGGATCGGCGAAGGCGTCGGCGACGCCG  
ACCTTGACGCGCCCGACCTCGGACGTCTGCGCGACCGGATCTCGATGTC  
25 GTAGACGGCGGCTGATCGAGCCCGGGCGAGAACGTACCTCGCGCGCGA  
15 TCGGGTCTGACGCGGCGCCCTCGGGGAGCGGGCCGACCTCGAACGCGTCG  
CCGGCGAGCGCGAGGCCGCTCGCGCACGTACCGGGAACGTACCGGTCTC  
30 CCCCTCGAGGAGCCAGTGCGGGCCCGCCCGACGGCTGSCAGCGCGAGC  
CCTCGGCGCTGGAGCCCCCGCGCTGGAGCCGGAC

25

30

20 Seq ID No 90 (&gt;Contig11)

35

GGCGACCCACATATCACATAGTAGAATCAGTGTGAGTTAGACAAATGTC  
GAGTGATGAGAAGGACAGAAGTGACAACTCTGTCTGATCACTGTAGAACGA  
GAGAGTATGAGCCTGCATACATGATAGCGGACATGAGAACGAGTGTANTA  
40 TGATGCTACTAAGAGAGTAACAGATCAGAGACTAGAGTAGAGCAATAGAA  
25 NTCAGAGATAAGTCAATGACGAGGAGTAGTGATAGAGCTCTTAATAATGG  
CTGAGGTCTGAAGATAGAAGTGATAGAGCGATAGATATACAATCGGTTGA  
AGCAGAGAGTAAGATAAGATCAGACACNGAGTACAGAGAGAGACGAATAG  
45 ATGGCGTGATNTCACAGAGAGGTGCGAGCGTAGCTGACGAGAGCAGAGAC  
GCAGAGTAAGTCACACCTAGATAGTTACGGCGAGAGACAAATGATAGGAA  
50 GGAGTGACGAGATCAACAGNCCGGAGCACAGAAGCGTGAGATGCGACCG

50

55

5

10

15

20

25

30

35

40

45

50

55

TGTAATAAACAGGAGACAAGAGCGACTACATAAGAGAGCGAAGCGAATAG  
ATAAGATATAAGCCCAGAGCAAAATAGAAGGAGAGAGAGAGTATTTGTAA  
TAAAGCAACAAGACGGAGAGAGCGAAGCAGCAGGCAACGATTAGAAGAAA  
GACGACAGGAAAGTGAAAGCGAAAGAGAGCAGGTAGAAAGAGAACCAAAA  
AAGCACGAAGGAAAAGGAAGCTTCTATGATAGGTGCGGGACAAGGCGTAG  
CTACAGGAGACAGCCGGCATAACGAGGAGCCGGTAAAAGCTAGCCTTTCAG  
AACACATCGGGAGCGCGTAAAGGCGGACCACGCTCGACGGGATCATGTAC  
GCCGACAGCGACGCCTTCAGCCCCGCGCGCACGTCCGGCGCGTCGCCCCG  
CGCGTCGCCGTGAGCAGCAGTAGGCGACCAGGCGGGCGTCGCCCCGGC  
CGTCTCGCGCAGGACCACGGCCGCCTGGCCCCACGCCGGGCACGCGCCGG  
ATCTGCGCCTCGACGTGCGCGAGCTCGATCCGGTGCCCCCGA3CTTGAT  
CTGGTGGTCCGAGCGGCCCTGGAACTCGAGCATCCCGTCGGGCAAGAAGC  
GCGCGACGTGCGCGGTCCGGTACATCCGCCCGCCGCGCGCGCGCGCAC  
GGGTGCGGCAGGAAGCGCTCCGCGGTGAGCCCGGGCTGCCCCACGTAGCC  
GCGCGCGAGCGGCGCGCCCCGCGATGTAAAGATCGCCGAGCGCGCCGATGG  
CGGGGCGGCGCATCGCGCCGTGAGCAGCAACACCTCGGCGTTGCGGACC  
GGCGCGCCGAGGGGGACCCACGTGACCCGCGGGTCGCTCGGCAGGACGCA  
GCCGGTCACCGCGATCGCGGCCTCGCTCGGCCCGTACATGTTGATGAGGT  
CGCCGTCTGTCTTCGCGTAGAAGCGCCGGACGAGATCGAGCGGCACCGCC  
TCGCCGCCCCACGAGGACCTTCCGCGAGGCTCGCGGGGAACGGCTGCTCGGG  
CCCCCGAGGAACGCCGCGAGCATCGAGGAGACGAAGTACGCGGTCTGTCG  
CCCCCTCGTCGCGCACGAGGCGCCGAAGGTAAGTCTCGGATCGCGGTGCCCCG  
CCGCCCCGGGCGACGACGATCCGCGCGCCGAACGAGAGGGGCCAGAAGAT  
CTCCAGACGGAGACGTGCAAGCCGAACGCGGCCTTGAGCAGGACCCGGT  
CGTCCGCGGTGAGCGCCAGTACCGCTGGATCCACTGCATCTGGTTGACG  
ATGGCGCGGTGGGAGATGAGGCTCCCTTCGGCGTGCCCGTCGATCCGGA  
CGTGTAGATGACGTACGCGCCGCTGTCCGGCGGCGGGCTACGGCGGGCC  
GCGCGTCGGAGCACGCGGCGATCTCGGCGGCCTCGGCGTCGAGGAGCAGC  
GTCGTCCAGCCCGCGGTGCGGAGCTCGTCGGCGATCGCGTCGTGCGTGAC  
GAGGAGGCGCGCCCGCGCGTCCCGCATCATGAAGGCGAGGCGCTCGCCGG

5

194

10

5

15

20

10

25

15

30

35

40

25

45

30

50

55

GGTACTCGTGGTCGAGCGGCAGGTAGGCGCCGCCACCTTGAGCACCGCC  
AGGGTCGCGACGACCATGTCTCGGAGCGCGGCACGCAGACCCCGACGAT  
CGTGTCGAGCCCGACGCCGCGGGCGCGCAGCAGCTCGCGAGCCGGTTCG  
CGCGCCGCTCGAGCTCGCCGTACGTGAGCGACTTGCCCTCGCTGCGCACC  
GCGACGACGTGCGGGTGCTGCTCGGCGCGCTCCTCGAACCACCGGTGCAG  
CGCGCAGGCCGACGGCAGCTCCATCGCGGGGCCGCGCAGCACGCCTCGA  
TCTCGGCGCGCTCGCCGGGGCCGACGTACTCGCCCTGGGCGACCGGACGC  
TCGGGTGGCGCGACAGGTCTCGAGCAGCGCCGCGAGGCGCTCGGCGAG  
GCGCTCGGCGTTCGCGCGCGCGCCCGGCCGACGCGTTCGTAGCGGAGCTCCA  
GCGACGCCGACGGGCCCGCGCCGCGCGCACTGCAGCCGGCAGGCGACCTGG  
TCAGACGTGCTCCAGACGTTCGAGCACGCGGGCCCGCGCGCCGTTCAGCGA  
CAGCGCCGGCGCGGCCTCCGCGGCGGAGAAGCCCCAGCTCATCCGGTGG  
CTCACCCCGGGCGCGCGCTCCTGGTTCGCGGCGCCCTCGGCCTCGGCGAG  
CGCGAGCCGCCGCGCAGCTCGGCGAGCGTGTCCGAGGCCGAGATCTCGA  
TCCGCACCGGCAGGAACCGCGGAACGGCCCCACCGCGCCCGCGAGCGCG  
TCCAGCGACCGCCCGTCGAAGCGGACGGCCACGGTGACCTCGGGCTCGTT  
GCCGCCGCTCATCCGCCACAGGAGCGACGCCACAGGGCCAGGAGCACGA  
TCCGCTGCGGGACCTGCCACGACGACGACCAGCGCTCGACCTGCGCCATC  
CCGCCTTGTCCAGATCGACCCGCGCGCGCCCCGAGCCGGCGCCGGCGCC  
GGCGCCCGCGCGGCTGAAGGCGAGGTGGAGCGGGGGCCGAAATGCGAGC  
GGCGCTCGGCCCAGAACCTGCGCCCGTTCGCCGGCGTCTCCGACTCGAGC  
ATCCCGTTGAGCCACTCGGCGACGTCCGCGTACTGCTGCTCGGGCGGCGC  
GCCCCGCGCCCGGTCGACGCGCAGAGCTCGCGGACGAGCGGGGCGATCG  
ACTCCTCGTCGACGCACACGCGGGCGCCGCGAGCACGAGCCGGCGCTCC  
TCCGGGCCGACGCGGACCAGGCCGACGCGCAGCCCGTTCGTCGCGCCGCG  
GTCTCCGAGAGGCGCGCGACGAGCCGCGACATCCGCTCGCCCTGCTCGG  
CTTCGGAGCACCCGACCCAGTCGTCTGTGACGCCCACGCGAAGCGCGG  
CTCGCCGACCACTGCGCGGCCTCGCCCGCCCGCTCGACGAGGCGCGT  
GCGCAAGATCTCGTGCCGCTCGGCCAGCGCGAGCGCCGCGCCGAGAGCC  
GTCCCTCGTCGCACGGGCGGTCACGGCGACGACGCCAGCGTCCGGCAC

5

195

10

5

15

20

10

25

15

30

35

40

25

45

30

50

55

CCGGGCGCCCCGCCTCCCGGTCGAGCGCGCGGATCGCCCGCTGCTGCGG  
CGAGAGGCTGAAGCCGGTCATCTCGTGGTCACTCATCCAGGTCGTCCTTC  
GGTGAGGTCTTCGCTTCGCCCCGGCGCGCCCGGGCGCGGAGGGTCACGGC  
GCJCCGGCGCGCGGAGGTCAGCTTGTCGAGCGCGGCGCCGCGCGCGCCT  
TCCGCTCGAGCTCCCGGCGGGCCGCCGCGCGCGCTCGAGCTCCCCCGG  
AGCTCCGAGACCGGGGTGTCCGGCCGCGCCGTGCGAGTGGCCAGGATCTG  
CCGTAATCGCTAAGGAAATTGTCGACCGTCGCCGCCCGGTACAGCTCGC  
TGCTGTGCTCGACGCCGAAGCGGAACGAGCCGCCGCCCTCGGCGACCGTG  
AGGACGAAGTCGAACGCCGTCTGTTGTCGCTCGCCCTCCAGCGCCTCGAG  
CTCGAGCCCCCTCGAGCTTCATCGGGGGACGTGCACGTTGCGCATGACGA  
ACTTCGCGTCGAAGAGGGGCACGTGCCCCGACGGCCCCCTTCGGCCGAGG  
GCCTCGACGAGCCGGTCGAACGGCAGGTCTTGGTGCTCGAACGCCTCGAG  
CGCGACGTCGCGCACGCGGCGGACCAGCGCGCCGAACGTGCGGTGCCCCC  
CGCAGTCGGTCCGAGCACGAGCTGGTTGACGAAGAAGCCGATCATCGGC  
TCGGTCTCGACCGGTTCCGGTTCGCGACGTCGGTGCCACGACGAGGTC  
CTCGAGCCCGGTGCGCTGGTGACGAGACGAGCTTGACGCGGCGAGCAGGG  
CCATGAAGGGGAGATCGCCTCCCGCTCGCAGAAGCCTTGATCTGGCGG  
GTGAGCTCGGCCCCGGCGTCGAGGCTCCGCCGCGCCCCGCGCCACGTCCT  
TCGCCCCGCCGGCTCGTGGTCGACCGGCACGCGGGCCCGGCGCAGCGCGC  
CCGAGAGCTTCGTCTCCAGTACCAGGCTCGCCCTCCAGGACCTCGCCG  
GACAGCCACGCCCGCTGGGCTGCGGCGAAGTCGACGTACTGCGCCGGGAG  
CTCCGGCAGCCGGAAGGCTGGCCCTGCGCGAAGCCGCCGTAGAGCGCGG  
CGAGCTCGCCGACGAAGACGCCGACCGACGAGCTCGAACACGACGTGG  
TGCACGACGAGCGCGATGACGTGCTCGTCTGTCGCTTCCGATGACCCG  
CACGCGGAGGAGCGGCCCGGGCTCAGGTCGAACGGCGCGAGGCTCTCCT  
CGAGGACGAGCGCCGAGACCGCCGCTCGAGGGCTCGCCCGGAGGTGC  
TCGAGGTCGGACATCCGGAACGGCACCCGGGCTCGGGCGCGACGACCGG  
GAACGGCACGCCGTCCCTGGCGCTGAACGTCGTCCGACGCGCCTCGTGGC  
GCCGCGGATCTCGAACAGGCTGCGGCGGAGCGCGTCGACGTCGAGCCGG  
CCCGTCGCGCGACACGAACGGCATGTTGTACGCCGGGCTGCCCGGCTC

5

10

15

20

25

30

35

40

45

50

55

GAGCTGATCGACGAACCACAGCCGGTGCTGCGCGAACGACAGCGGGAGCG  
GGCCGTGCGGGGGATCCGCGCGATCGGGGGAACTCGCGCCGCGCGCC  
TCGCGCGGGCGCGGGCGTCGACCTGGGCCGCGAGCGCCGCGACCGTCGG  
CCCCTGGAAGAGCGCCCGACGCGGGAGCTCGACGCCGAGCTGCGCGCGGA  
5 TCCGGGACATCACCTGGGTGCGGACGAGCGAGTCGCCGTGCAGGCCGAAG  
AAGTCGTCGTGGACGCCGATCTCGTGGACGCCGAGGAGGGCGCTCCCAGA  
15 TCGCGGGCGATCGCGCGCTCGGACTCGGTGCGACGCGCGGGCGAACGCGGC  
CCCGGCGTGCGCGGGGAGACCGCCGACGTGCGGAGGGCGTTTCGGCGCGG  
GCGCGAGGGGGCGGTGCGCCGCGCCGCTCGGCGCGGGGGCGCCGCGG  
20 CGCGGCGCGATCCAGTGCCGCGCCCGCTCGAACGGGTACGTGCGCAAGCG  
10 GACGAGCGCGCCGGGGACGCCCCGCGGACGGGCCGTCCAGTCGACGGC  
GTGGCCCGCCTCCAGAGCTGGCCGAGGGCCTCGGCCAGGCTCGCGGGCT  
CGGACGCGGCGTGGGTGCGACCCGAGGCTCGCGATCGCGGCGCCGCCGCG  
25 CCGGCCAGCGTCTGCCGCAACAGCGTGGTCAGCCCGCGGCCGGGGCCGAC  
15 CTGAGGAACAGGGCGTGCCCGGACGCGAAGAGCGCCTCGACGCCGTGCG  
TGAAGCGGACCGGCTGGCGGAGGTGCCGCGCCAGTAGGCCGGATCGGTG  
30 GCCTCGGCGTCTGGTGAAGAGGGCGCCGGTGACGTTGAGACCACGGGGAT  
CTCCGGCGGGGAGAGCCGCGCGCGCCGACGCTCTGAGGAACGGGGCCA  
CCGCGCCGTGATGAGCGCGCAGTGGAACGCGTGGGACGTCTGCAGCGGC  
20 CGGGCGAACACCTCGCGCGCCTCGAGGCGCGCGCGAGATCGCGGATCGC  
35 GCTCGCCGGGCCCCGCAACCGTGAGCTTCGGGCTGTTGACCGCGGCGA  
TCTCCAGGCCGGCCTCGAGGAGGCCCTCGACGTCCGCGGCCGCGAGGCCG  
ACGGCCAGCATGCTCCCGGCCGGCGCCGCTGCATGAAGCGCCCCGATC  
40 GATGACCAGGGACATCGCGTCCTCGAGCGTGAACACGCCCGCGACGCAGG  
25 CCGCCACGAGCTCGCCGAGGCTGTGGCCGATCATCGCCGCGGGCTCGATC  
CCCCAGCTCATCCAGAGCCTGGCGAGCGCGAGCTCGACGGCGAAGAGCGC  
GGGCTGCGCCAGCGCGGTGCCGAGCAGCGTGCGCCCGTCGCCCTCGCCCT  
45 CGCGGAAGACGACCTCGCCGAGATCGAGCCGCGCGCCCGCGCCCGCC  
GCGCACGCGTTGAAGGCGCTCCGGAACGCCGCCTCCTGCGCGTAGAGCGC  
30 GCGGGCCATCCCGACGGCCTGCGCGCCCTGGCCCGGAACGCGAAGACGG

5

10

15

20

25

30

35

40

45

50

55

GCGCGGCTCATCGGGGCGCGGAGCGCGCTCGCCCCCTCGCGGGCGAGCC  
CCTGGATCGCCTCGGCGCGCGTCCGGGCGACGACCGCCCGGCGGTACGGG  
TGCTCCGCGCGCCCGGTCTGGAGGGTGAACGCGACGTCGTCGAGCGGGAC  
GTCGGTCGCCTCGAGGTGCGGGCGAGCTGCGCGCAGGCCGTCGACAGCG  
5 CCTCCGGCGTGCGCGCCGAGAGCGTCAGCACGTGATCGCGCTCCGGGGCC  
GGGCGCGGGGCGGCGAGCGGGGCGGCTCCTCGAGCACGACGTGCGCGTT  
CGTCCCGCGGATCCCGAACGAGCTCACGCCCGCGCGCGGGCGGAGCT  
CGCGGGGCCAGGGCGCGCCTCCCGCGGACGAAGAACGGGCTCGCCGCG  
AGGTCGAGCTTGGGGTTCGGCGCCTCGAAATGGACGCGAGGGCGGGATCTC  
10 GCGGCTCCGCGACGACGTGCGCGCCTTGATGAGGCCCGCGACGCCCGCG  
CGGCGTCGAGGTGGCCGATGTTTCGCCTTGATCGAGCCGAGCGCGCAGTAC  
GCCTTCCTCGGGGTCTTGCGGCGGAAGGCCTGCGTGAGCGCCTCGACCTC  
GATCGGATCGCCGATCGCGGTGCGGTGCCGTGGGCCTCGACGTAGCCGA  
25 TCGAGCCGGGATCGACGCCGCGGCGACCGACTGCGCCTCGGAGATCGCCGCC  
GCCTGGCCGTCGACGCTGGGCGCCATGAAGCCGACCTTGCGCCCGCCGTC  
GTTGTTGACGGCGGAGCCCCGGATCACCGCGTGGACCGTGTTTCGGTCGC  
GGAGGGCGTCCGCGAGGCGCTTCAGCGCGACGATGCCGACGCCGCTGCCG  
30 CCCACGGTCCCTCGGCGCGCGCTCGAACGGCCGGCAGCGGCCGTCGGG  
GGAGCAGATGCTGCCGGGCACGTACGGATAACCGCGCTTCTGCGGGATGC  
CGATGGAGACGCCCGCCGAGCGGAGATCGCACTGGCCGCCGAGGAGG  
20 CTCTCGCACGCCATGTGGACGGCCACCAGCGACGTGAGCACGCGGTCTG  
CACGACCACGCTCGGCCCGTGGAGGTGAGTTTGTACGAGACCCGCGTCG  
CGAGGTAATCCTTCTCGCTCGCCAGCATGAGCGCGTGCGGATCGACGGTG  
40 GCGCGAGATCCGGGTGCGAGAGGAGCTGGAGGAGGTACGTGTTGGAGCC  
GCACCCCCGAAGACGCCGATCGCGCCCGGAACCGGGCCGGATCGCAGC  
25 CGGCGTCCTCCAGGGCGGCGACCGCGCACTCCAGGAAGAGGCGTGCTGC  
GGGTCCATGAGCTGCGCCTCGCGCGGCGAGTACCCGAAATAGGACGCGTC  
45 GAAGCGGTGATGTCGTGAGCAGGCCGCCCGCGCAGACGACGCGCGCCC  
CGGGGGCGCGCTCGCGCCGACCGCGGCTCCTCGCGCTCGCTCTCCGGG  
30 AAGCGCGGATCGACTCGACGCCGCGCCGACGTTCTCCAGAGGGCGTC

5

198

10

15

20

25

30

35

40

45

50

55

GACGCTCGGGGCGCCGGGGAAGCGGCCCGCCATGCCGACGATCGCGATGT  
CGTCCCCCGTCCCTCGGTCTCGATCGGCTCTGACATGGCTATCCTCGCC  
CCCGGCGGCGTTCGCGCGTTCGCGGCGCGCCTCGGCGCGCTGCGCCCCGACG  
TCGGCCGGCTCGGCCTTGACCGTTCGCCGCGTTCGAGCCGCTGCGCCAGTTG  
5 CTGATGGTTCGGGTACTGGAACAGGTTCGGTCAGCGACACGGCCTGCGCCG  
CGGCGCCCTCGTTCGGGCGCGCGCGCCGCGATGCGCTCGGCGAGCAGGCGC  
15 TGCGCGCGCACGAGGAGCAGCGAGGTGAAGCCGAGCTCGAAGAGGTTGTC  
GGTCACGCCGACGGCCTCGACCTGCAAGACCTCCGCGAGCACCGAGGCGA  
TGAGCCGCTCGGTTCGCGGTTCGCGGGGCGACGCGCGCGCGCGCGCGCG  
20 ACCGCGCGCGGATCCGGCAGGGCGCGCGGTCCACCTTGCCGTTTCGCGCT  
10 CAGCGGCAGCGCCGGGAGGACGACGACCTCCGCGGGATCATGTACTCCG  
GCAGCTTCTTCGGACGAAGTTCGCGGAGCGCGCGCCATCGCCGTTCGGCG  
CCGACGACGTACGCGACCAAGCGCTTCTCGCCCGACGGATCGGTCTTCGC  
25 CGCCACGACCGCCTGCTCGACCGAGGGTGCTGCGCGAGGGCGGCCTCGA  
CGCCCGGAGCTCGATGCGGAAGCCGCGGATCTTCACCTGATGGTTCGGTG  
15 CGCCCGAGCAGCTCGATGGTCCCGTTCGGCGAAGTAGCGGCCAGGTTCGCC  
TGTCCTGTACAGCCGCTCGCCGGTTCGTGGGGTGCTTCAGGAACCGTCCC  
30 GGGTCCGCGCCTCGTTCGCGCCAGTATCCGAGCGCGACGCGGATCCCGCCG  
ATGTGGATCTCGCCGGGACCCCGATCGGACACGGCTCCAGCCCCCTCGTC  
20 GAGCACGTAGGTGTGCTGGTTTCGCGAGCGGGCGGCCGTAGGGGATGCTGC  
35 GCCACGCCGGGTTCGACGTCCGCGATCGGTTGGGCGATCGACCAGATCGAC  
GCCTCGGTTCGCGCCGCCAGGCTCACGACGCGGGGCGCGCGCAGGCCGC  
GCGGATGCGATCGGGGAGCTTCAGCGGGATCCAGTTCGCCGCTCATCATGA  
40 CGAGGCGGAGCGACGACAGCGCCGGTTCGCCCCGCGCCGGGGACGCGTCC  
25 ATGAGCATCTCCATCAGCGCCGGGACCGAGTTCACACGGTCAACCGCTC  
GCGCTCCACGAGCTCGCGCCAGTGCCCCGATCCGAGCGCGGGTACGGT  
CGGGGATCACGACGGCGCCTCCGGCGGCGAGCGTCCCGAACACGTCGTAG  
45 ACCGACAGGTTCGAAGCTCAGCGACGAGAGCGCGAGACCCGGTCTCCGG  
GCCGACGTTCGAAGCGGCGGTTGATGTCGAGGACCGTTCACCGCGCCGC  
30 GGTGGTTCGATCATCACGCCCTTGCGCAGCCCCGTGGACCCGGACGTGTAG

5

10

15

20

25

30

35

40

45

50

55

ATCACGTAGGCCAGGTCGTCCGTGCTTCCGCCGGGCGGCCGGCGCGAC  
GGGCTGCTCGCGCCACCGCTCGTCCGCGTCGACGGCGAGGCGCTCGATGC  
CCGCGGGCCAGGCGATCGTCCCGTCGACCGCCGACTGCGTGAGGACGAGG  
CGGACCTCGGCGTGCTCCAGGAGGTGCCTGAGGCGCTCCTCGGGGAGGCG  
5 AGGGTCCAGGGGCAGGTAGGCGGCGCGGCGCGCAGCACGCCGAGCACGG  
CGGCCACCTGCTCCCAGCCCTTCTCCATGACCACGGCGACGAGCGCGTTC  
GCGGTGCTCCGGAGCGCGAGGCCGCGCGCGGATCGCCTCGGCGCGCCG  
GGCGAGCTCCCCGTAGGTGAGGCGCCGCTCGGCGTCGACGACCGCGCACG  
CGTCGGGTGCTCGACGGCGCGCTCAAAGAACGGCTCCTCCAGCCGGAGG  
10 TGATCCGGGGTTGCGACCGCGGTGTCGTTCCACGCGACGAGGGCGCGCTC  
GCGGTCTTCCGGCGCGACGGAGAGCGCGCGGACGCGCTGCGCGGGTCTCT  
GAGTGGCGCGGAGAGCACGCTCTGCATCGTGGCGAGCATCCGGTCGATG  
GTCCGCCGCTCGAAGAGGTGCGACGTTGTACTGGAGCGAGATCACGTCGCG  
25 CCCGCCGCGGGCTCGACGCTGAAGCGCAGGTGGAAGCGGTGGCCTCGA  
CCGGGAGATCGAGCGGCTCGATCCGCACCTCGCCGAGCTCGAGCGCCTCG  
15 GTTGGGGCGTTCTGCACGACGAGCATGACCTGGAACAGCGGCGAGCGGCT  
CAGGTGCGGGCGGGGGTTGACCGCTCGACCACCTTCTCGAACGGGGCGT  
CCTGGTGCTCGAACGCCTCGAGCGCGACCTTCCGCGCCCGCGAGAGGAGC  
TCCTCGAAGGTGCGGTGCGCGCCGAGGTGAGGCGCATGACGATCGTGTT  
20 CACGAAGAAGCCGACGAGGGGCTCGAGCTCGGGGCGAGGCCGTTGGCGA  
35 CCGCGGTCCCGATGGCGAGGTGCTCCTGGCCCCGAGCTGCGCCGAGGAGC  
ACGCCGAGGGCGGCGAGCAGGACCATGAAGCGGTGGCGCCGCGGCTCCG  
GGCGAGCTCGTCGAGCTGCGCCACGAGGCGCGCTCGAGCGGGAGGACCC  
40 GCTCCGCGCCGCGGAACGTCTGGACGGGCGGCGCGGTGATCGGTCTGG  
25 AGCTCCAGGACCGGCAGCCCGCGGAGGGTCTGCTGTCCAGTGAGCGAGCTT  
GTCCGGCAGCCGCTTCCCCGCGAGGTGGCGGCGCTGCCACACCGCGAAAT  
CGACGTAAGTGGAGCGGCGAGCTCGGGCATGTCCGCGGGCCCGCGCCCCGC  
45 GCGCGCCGGTAGAGCTCCGCGAGATCGCGGACGAGGGGTTGGAAGGACCA  
GGCGTCCGTGACGATGTGGTGCTGGACAGGACAGGACGACGACGTCGT  
30 GGTGAGGCGGAACAGCCTGGCGCGGAACACGGGCCCGCGCGAGGTCC



5

200

10

15

20

25

30

35

40

45

50

55

AACCCCGTGGCCTGCTCGCGGACGCCCAGGCGCGCGCCGCGGCCTCCGC  
CTCGTCCGGGGCGTGCCGCGGAGGTCCGACCACCTCCGCGGGGGCCGCT  
CGGGCTCGCAGATCTTCTGCGCCGGCGTGGGGCTCGCGACGAACACCGTG  
CGCAGGCTCCAGTGCCCGCCGACGAGCGCGGAGCGCGGAGGACAGCGC  
5 GTCGACGTCGACGAGGTTCCGCGAGGCGGACCGCTGCACCACGTTGTAGG  
CGGTCCCGCCGGGAGCAGCTGCTCGAGGACCCACAGGCGCTCTTGCTCG  
TACGAGAGCGGATACGGCTCGTCCGCCGGCGCGCGGCCAGCGAGGGCGC  
GATCTCGCTGGCGGGCACCGTCGCTGCGGCGGCGGTGGTCGAGGCGGCGC  
CGGAGGAGAGGCGATCGGCGAGCTGGTGGAGGGTTGGGTGCTCGAAGAGC  
20 GTGCGGAGGGTGGTGCGGATGCCGAGGAGGACTCGATGCGTCCGAGGAC  
CTGCATGGCGAGCAGGGAGTGGCCGCCGAGGTCTGAAGAAGCTGTCTGTCTC  
GTCCGACGCGGTCCAGGTGGAGGACGGATTGCCAGATGTGCGCGAGCTCC  
CGCTCGAGCTCGCCCCAAGGGGGCTCGTAGTCGGCGTGC CGCGGCGGTGG  
25 CGCAGGGAGGAGCTTCTTGTGACCTTGCCCCGAGAGGGACATGGGCAAGG  
CGGGGAGCAGGACGAAGTGGGCGGGCACCAAGGCGTCGGGCACCAGGCGG  
GCCATGCCCTCGCGCAGGTCCGCGCTCGGAGGGCGGGTCGGCGCCCGGCAC  
30 GACATAGGCAATCAGGCGCGCGGCGCTGCCTTGCCCGTGAGGACGACGA  
CGCCCTCGCGGACGGCGGGCAAGCGTCCGAGGGCGGATTCCACCTCGCCG  
AGCTCGACACGGCGACCGCGGAGCTTGACCTGCTCGTCGCGGCGTCCGGC  
20 GAAGGCCAGCTGTCCGTGCGGGCGCCAGCGCACCAGGTCGCCCGTGCGGT  
AGAGGCCTGCGCCGGGCTGGCCCAAGGGATCGGGCAGGAAGCGCTCTGCG  
GTCAGGTCCGTGCGTGTGTAGCCCTGGGCGAGGCACGCTCCGCCGATGTA  
CAGCTCGCCGAGGACGCCGGGCGGGACGGGCTGCATGTGCGGGTCGAGGA  
40 CGTAGACGAGGGCGCTGTCCGATGGGTCCGCCGAGCGGGGGCTCGTCGCCG  
25 AGGTCCGCGACCTCGGCCGACGGTGGTGATGACGGTGGCCTCGGTGGGGCC  
GTACATGTTGAAGAGGCGGAAAGGGAGCGGTCCCGGAGCGGATGGAGCT  
TGTCGCCCGCGACGGTCATCGCGCGCAGGGCGATGCCGGTCCAGTCTTGC  
45 TCGAAGCACGCCTCGGCCAGGGGCGTGCCCATGAACGAGAGCGTGGCCCG  
CTGAGCGACAAGCCAGGAGACGAGCGCTGTGGGAGAGCGGAGCGCGTCTGT  
30 CGTCGGCGAGGAGGAGTGACGCGCCGAGGCGAGCGGCGTCCAGATCTCG

5

201

10

15

20

25

30

35

40

45

50

55

TAGACGGAGCGCTCGAAGCCGCTGGAGGCCAGCTGAGTCCAGCGATCGCG  
GGGTGAGAGCGCGAGCAGGTGCTGGAAGAAGGAGACGAGCCTTGAAAGGC  
TCGCATGGCGCACACAGACGCCCTTCGGCGTGCCGGAAGAGCCGGAGGTG  
AAGAGGACATAGGCCAGGTCTGTCGGGCTTGAGACGAGAGGAA7GTGGGT  
5 GCTGGGCGCGCACGCCCCGTCCTGGACGAGGTGGACGGGGCAGGGGGCGG  
CGGTGAGCTTGTGGCTGGCCTGGCTGCTGGTGAGCACGAGCGGGCGCGG  
CAGTCGGCGAGCATCTCGGCCAGGCGCGCCGGGGGTTGGCGGGTTCGAG  
CGAGGCATAGGCGGCGCCTGCCTTGAGGACGGCGAGCTGGGCGGCGACCA  
TGCGGGGCGAGCGCTCGATGCAGACGCCGACGAGCTGCCGGGGCCGACG  
10 CCGCGGTGCGCGAGCCACAGGGCGAGCTCGCTGGACCAGGTGCTGAGCTC  
TGCGTAGGTGAAGCGCTGGTGTCCGAACTCGAGCGCCGTGGCGTCCGGCT  
GTCGAGCGGCGTGGGCCTCGAAGAGCGCATGGACGCAGGCGGGGGCCGGG  
GCGGAGGCGGCCTGTCTGTCGGCGGCGAGCGCCGCTCCAGTCTGTCAGGAG  
CAATGCGCGCTCGGCGTCGGAGAGCATCCGGAGCTCGGAGAGCGGTTCGAC  
15 CGGGGTGCTCGACGGCGCTTTCGAGCAGGAGCACGAAGTGGCGGCCATC  
CGCTCGATGGTGGCGGGTTCGAAGAGCTGCTGGTCTGTAAGCGCAG  
GGCGATGCCGGAGTCGAGCTCTGCGGCGAACAAGGCGAGATCGAACTCGG  
CCGCTGCCTGCTCTGTCGGCGAGCGTGGTGAGCTCGAGCTCTCCCTGCGCG  
ATCCGCACGTCTCCGCGCCGGTGGTGAGCGCGGCGAGGCGGGGATCCAG  
20 CGACGGCAGAGCGCCCTGGAAGGCGAAGGCGACGTGGAAGAGCGCGCCGC  
CTCGCCGGGCGCGCCCCGGGGCTCTGCGAGCAGGTGCTGGAGGGCGCTG  
TCGCCGTGGGCCAGCCCGTCGAGGAACGCGTCGCGCACGCGGGCGACGAG  
CGCGTTCGAAGGACGCGGCCCCGCGCAGCGCCACGCGCACGGGGAGCATCT  
GGACGAAATAGCCGAAAGCCCCGAGTGCTCTCGTCTGTTCCGCCCCGCC  
25 GAGGGGACGCCCACGACAAGGTCTTCTGCCCCGCTCGCGGATGGAGCAA  
GACGGTGAGCGCCGACAGCAGGACCGAGAAGAGCGTGGTCCCGCGCTCGC  
GCGCGAGGCGCGCCAGCGCTCCGGTCAGGGGCTTTGGCAGCGTGATCGCG  
TGAGCGCGACCGCGGCGAGGGCTCGCGTCTGTCGGGGCCCGTTCGGGGG  
AAGGTTCATGGCGGTCTGTCGCGCCGTCGAGCGCCTTGCGCCAGTATTCTG  
30 CTCCGCGGGCGCCTCCCGCGGCGAGGGACAGCTACGCGGCGGCGGAAG

5

10

15

20

25

30

35

40

45

50

55

AAGCTCGACGGCGGGCGGCAGCTGCGGGGGCCGGCCCGCGCGCAGCGCCGA  
GTACAGCTCCCCAGCTCGCGAACGAGCAGCGCGAACGACCAGTAGTCGA  
CCACGAGGTGGTGAACGACCACCGTGAGCAGCGGCGGCTGCCCTCTCCG  
CGCCGCCAGACATGCACCCGAGCAGCGGTCCGCGCTCCAGGTGGAACGC  
5 GCGGGCGGCACCTCGTCCGCGGGGCGACGATCTCGCGCTCGTCCAGCG  
CCATCGCCGGCTCTTCGGCCCATTCAGGGCGACATGGCGGTGGACCTGC  
15 TGCAGCGGATGGCCGTGCGCGGTGAGGAACGTGCGGAGCGCCTCGTG  
CCGCTCGACGAGGCCCTCGAACGCGCGGCGCAGCGCGGCCACGTGACGC  
CGGCACCGAGCCGACCGTCTGCCCCAGGTTGTAGAGCGCGCGTCCGCC  
20 GACTTCTGGCACTCCAGCCACATCGCCCGTGCCTTCGGTCAGCGCAAA  
CGGCTCTTCGGCGTCCGACCCGCGGCGACGAGGGGCCGTCCGGACGGG  
TGAGCCCTCGCTCCAGGGCCGTGCTGCGGCGGCGGAGGTGAGGCGGCG  
CCGGAGGAGAGATGACTGCGGAGCTGCGCGAGGGTTGGGTGCTCGAAGAG  
25 CGTGCGGAGGGTGGTGGCGATGCCGAGGGAGGACTCGATGCGTCCGAGGA  
CCTGCATGGCGAGCAGGGAGTGGCCGCCGAGGTGGAAGAAGCTGTCGTGT  
15 CGTCCGACCGGTCGAGGTGGAGGACGGATTGCCAGATGTGGGCGAGCTC  
GAGCTCGAGCTCGCCCCAGGGGGGCTCGTAGTCGGCGTGCAGGCGGGGG  
30 GCGCAGGGAGGAGCTTCTGTGCGACCTTGCCCCGAGAGGACATGGGCAAG  
GCGGGGAGCAGGACGAAGTGGGCGGGCACCAGGGCGTCGGGCACCAGGCG  
GGCCATGCCTTCGCGCAGGTGCGCGTCCGAGGGCGGGTGGGCGTCTGGCA  
35 CGACATGGGCAATCAGGTGCGCGGCGCTGCCTTGGCCGTGGAGGACGACG  
ATGCCCTCGCGGACGCCGGCAAGCGTCGCAGGACGGATTGACCTCGCC  
GAGCTCGACGCGGCGACCGCGGAGCTTGACCTGCTCGTCGCGGCGCCCCG  
40 CGAAGGCGAGCTGTCCGTGCGGGCGCCAGCGCACCAGGTGCGCGGTGCGG  
25 TAGAGGCGTGCGCCGGGCTGGCCGAAGGGATCGGGCAGGAAGCGCTCTGC  
GGTCAGGTCCGTGCGTGTGTAGCCCTGGGCGAGGCACGCTCCGCCGATGT  
ACAGCTCGCCGAGGGCGCCGGGCGGGACGGGCTGCATGTGCGGGTCGAGG  
45 ACGTAGACGAGGGCGCTGTCGACGGTTCGCGCGAGCGGGGCTCGGCGCC  
GAGGTGCGCGATCTCGGCGACCGTGGTGATGACGGTGGCCTCGGTGGGCC  
30 CGTACATGTTGAAGAGGCGGAAAGGAGCGGGCGCCGAGCGGATGGAGC

5

203

10

5

15

20

10

25

15

30

35

40

25

45

50

55

TTGTCGCCGCCGACGGTCATCGCGCGCAGGGCGGAGCCGGTCCA3TCTTG  
CTCGAAGCACGCCTCGGCCAGGGGCGTGGGCATGAATGAGAGTGTGGCCC  
GCTGAGCGACAAGCCATGAGACGAGCGCCGTGGGAGAGCGGAGCGCGTCG  
TCGTCGGCGAGGAGGAGGGCAGCGCCGAGGCGAGCGGCGTCCAGATCTC  
GTAGACGGAGGCGTCGAAGCCGCTGGAGGCAACCTGAGTCCAGCGGTCGC  
TGGGCGAGAGATCGAGTCGGAGGTGGAGGAAGGAGACGAGCCTTGAAAGG  
CTCGCATGGCGCACACAGACGCCCTTGGGGGTGCCGGTGGAGCCGGAGGT  
GAAGAGGACATAGGCCAGGTCTGTCGGGCCTGGAGACGAGAGGAATGTGGG  
TGCTGGGCGCGCACGCCCCGTCCTGGACGAGGTGGACGGGGCAGGGGGCG  
GCGGTGAGCTTGTGGCTGGCCTGGCTGCTGGTGAGCGCGAGCGAGGCGCG  
GCAGTCGGCGAGCATCTCGGCCAGGCGTGCCGGGGGGTTGGCGGGGTGCA  
GCGAGGCATAGGCGGCGCCTGCCTTGAGGACGGCGAGCTGGGCGGCGACC  
ATGCGGGGCGAGCGCTCGATGCAGACGCCGACGACGCTGCCGGGGCCGAC  
GCCGCGGTGCGCGAGCCACAGGGCGAGCTCGGTGGACCAGGTGCTGAGCT  
GTSCGTAGGTGAAGCGCTGGTGGCCAACTCGAGCGCGGTGGCGTCCGGC  
TGTCGAGCGGCGTGGGCCTCGAACAGCGCGTGGACGCGAGGCGGGGGCCGG  
GGCGGAGGCGGCCTGTCTGTCGGCGGCGAGCGCCGCTCCAGTCTGTCGAGGA  
GCAATGCGCGCTCGGCGTGGAGAGCATCCGGAGCTCGGAGAGCGGTGCA  
CCGGGGTGCTCGACGGCGCTTTCGAGCAGGACCACGAAGTGGCGCGCCAT  
CCGCTCGATGGTGGCGGGGTGCAAGAGCTGCTGGTCTGTAAGCGCA  
GGGCGATGCCGGCGTCTGAGCTCTGCGGCGAACAAGGCGAGATCGAACTCG  
GCCGCTGCCTGCTCGTGGCGAGCGTGGTGAGCTCGAGCTCTCCCTGCGC  
GATCCGCACGTCCCCACGCCGATCGCGAGGGCTGACAGGCGTGCATCCA  
GCGATGGCGGGGTGCTCTGGAAGGCGAAGGCGACGTGCAACAGCGGTCT  
CGCTGCGCCTCGCCCTGCGCTCGCGCGAGCAGGTGCCGAGGGCGCTGTC  
GCCGTGGGCCAGCGCGTCTGAGGAACGCATCTCGCACGCGGGCGACGAGCG  
CGTCAAGSACGCGGCCCCGCGCAGCGCCACGCGCACGGGGAGCATCTGG  
ACGAAGTAGCCAAAGGCCCTGGCGCTCTCGTCGTCGTGCCGCCCCGCCA  
GGGACGCCACGACAAGGTGCTCTGTCCGCTGGCGCGATGGAGCAAGA  
CGGTGAGCGCCGACAGCAGGACCGAGAAGAGCGTGGTCCCGCGCTCGCGC

5

10

15

20

25

30

35

5 GCTCGACGGCGGGCGGCAGCTGCGGGGGCCGGCCCGCGCGCAGCGCCGAGT  
10 GCGAGGCGCGCCAGCGCTCCGGTCAGGGGCTTTGGCAGCGTGATCGCGTG  
AGCGCGACCGCGGCGAGCGCCCGCGTCGTGGCGAGCCCGGTGCGGGGGA  
GGTCGATGGCGGTTCGTGCGCCGTCGAGCGCCTTGCGCCAGTATTCTGCT  
CCGCCGGCCGCTCCCGCGGCGAGGGACAGCTCACGCCGGCGGCGAAGAA  
15 ACAGCTCCCCAGCTCGCGAACGATCAGTGCGAACGACCAGTATTGACC  
ACTACGTGGTGATCCACCACCGTCAGCACTCTGCGTTTAGTCCTTTCTCC  
TGTTGCGCCTATTAATTGCTACTATGGATCCACACTGCTCCGCCTCTTGT  
ATCTCCCTTATCTGCACTTGCTGCGCTTNACCTTTACGTTCTCTCCCTG  
20 CTCTATACTATTTCTTCCCGCGTTCTCGTCTTATCTGCATTTGTCATA  
TCGTATCTTCATATACCTTTCTTTGCTATCCTTACTGCTTCTCGACCTT  
ATGTGCGTCTGTCTTCCCTTTCTNTATTATTTCTCTGTCTCACCGCTCTN  
TGCTCTGTCGCTCCTATCACTAAATTATGTCTCTATCACTGCTACTATCT  
25 GAAGCTGATCTTCGAGATCTCGCTNGGTGTCACTCTTTATCTCATAGNCG  
15 CCTCTGTCTTCTTGTCTCCTTAAGNCTGATTTTCTCGCTCTATTCTGAC  
TACTCTGCTGTCTCTCACATACGTGTTCTTGAATCGTATTGCGGTTCTCG  
CTACTGTGATATCCATTGCGGACCTCTACTGCTCNTCTNTATGCTATACT  
TCTTAGTCTCTTACTACGTTNGTCTGATATNTTGCTGACGACGTCATGTC  
ACGCTCGCAACTCTTCANTTCTATCGTATACGCTGATCATATTTCTGT  
20 GAGGCTGATGTACTATACGTAATTACCTGTATACGTCGTCTATCTACTCT  
35 CGTGCTTCACTCTTTCTACTCC

Seq ID No 91 (&gt;Contig12)

40

45

50

40 CCCCCCGCGTCCGCCGGTACGTGCGGACCGCCGCCCGAGCAGTCCC  
25 CGCGCTCGCGCCGGAGGAGCGGGAGGCCGCGCGCCGCTGTGCGCCC  
TCGGCGCGGCGCCGCCGAGGTCCGGCGCCGCGGGCTGACGCGGGCGCCG  
CTCTCGTACGGGCGAGCCGCATCTACTTCCTCGAGCAGCTCTCGCCCGG  
45 CAAGCCGCTCTTCAACGTCCCGGGCGCGGTCCGGCTCCGGGGCCCGGTG  
ACGTCGCGCCGCTCTCGGCAGCGTTCGGCGAGATCGTGCGGCGCCACGAC  
20 GCCCTCCGCACGTGATCGCCAACGTGACGGCGAGTCCTGCAGATCGC

55

5

205

10

5

15

20

10

25

15

30

35

40

25

45

30

50

55

GCAGCCGCACCGGGCTTCGCGCTCGACGTGGTGACCTCGACGCCCGAGG  
AGGCGGCCGAGCTCGACCGCGGGCTGCGCGCCGAGGCGTGGCGGCCCTTC  
GCGATCGGCGCGCCGCGCTCCTGCGCGCCACGCTGTTCCGCCTCGCGGA  
GGACGAGCACGTGCTCCTCGTCACGATGCACCACGTGGTGTGCGACGACT  
GGTCGCTCGGCGTGATCCTCCGCGAGCTCCTCGCGCTGTACGCGGGCCGC  
TCGCTCCCGCCGCGCGGCTCCAGGTCAGCGACTTCGCGGCGTGGCAGCG  
CGAGATGGTCGAGTCGGGGGCGCTCGACGGCCAGCGCGCTACTGGCGAG  
AGCGCTCCCGGGGCTGTCCCGGGCGAGCATCTCGGCCGGCGGGGGCG  
GAGGCGCCGAGCCACGACCCGTCGCGCGCCATCGAGGAGATCGCGCTCTC  
GCCGACAAGGCGGCGGCTCGAGGCGCTCGCGCGGGAGGGAGCGA  
CCCTGTTTCATGGTGCTCCTCGCGCTCCTCGACCTCGTGATCCATGCGCGG  
TCCGGCGCACTGGACATCGCCGTGCGGACGCCCATCGCCAACCGGAACCG  
CCCGAGCTCGAGGACGTGGTCGCTCTTTGACGAACACGCTCGTGATCC  
GCGTCGATCTCGCGCGCGCGGGGCGTTCCGCGACGTGCTCGCGCGGGCG  
CGCGTCCAGGCGCTCGACGCCCTTCGCGAACCAGGACATCCCGTTGATGT  
CGTACCCAGGATCTGAAGCAGGAGCGCGACCACGCGCAGCACCCGCTCT  
TCCCGCTCTGGTGGCGCTCCAGAACCGCGCGAAGCCCGCGTGGAGGTC  
CGCGGGCTCCGGTTCGAGCCCTGCCCTCCGGCCGAGCTCGTGCACTT  
CGAGGTGCGCCCTCCTGCTCTGGCCGGCGGACGACGGATCGGTCTGGGGC  
ACTTCGAGTTCGCGCGGATCGCGTTCGACGAGGGCGCGCGCAAGGAGATC  
GCGGCCGCATTACGCACCTCGTCGACGCGGTGATCGCCGGCCGACGC  
GCCGGTGTGACGCTCGTGAGGGCGCCCGCGCCGAGGCCGCGGAGCGC  
AGGCCGCGCTCGGCGAGGCGTTCGCCAGGGCGGCGACGGCGCGCCTCGGC  
CAGCTGCGGCGTCTGCTCGGCGGGCGACCGGACGCCCCGCGAGTAGCGGTG  
AGCCCTCGGCGGCGGCCAGGCGCACGCGGAACGGCGAGGCTAGCCGTGG  
ACGCGCGGCATGGGGTCGATCGCGCTGGGGACGCGGCCCGCAGCAGCTG  
CTTGATGGCGAGCGAGATGTGCAGGATGGCCACGTAATTGCCGTGGCAG  
TATGGATCCCTGACTCCCAGAACAGTAGTTTCTGCTCGGCAGGCGCCG  
GGCTGAAGTGTCCGGGGCGTCGATGTGCTCGTGGTCTGTCATCGCCGA  
GGCGCTGCAGGCCATCACCAGCGCGCGGGGACCTTCTCCTCGTGCC

5

206

10

15

20

25

30

35

40

45

50

55

GCGTGCCACGCCCCGACCGTG TAGTCGCGCACGCAGAGGCTCGTGACGCCG  
GTCGACGGGGGACGGAAGCGCAGCGCCTCCAGCACATAGCCGGTGATGGC  
GGCGTCGTCTCTCGACGTTACCCAGTTGAGCGCGTCGCGCAGGACGCGCG  
GGCGCTTCATCAGCTCGACCAGGGCGTTGACGATCGCGCCGCCGCTGAGA  
5 TCCACGCAGCCCATGAGCAGCCCCAGGATCACGTCGCGGATCCCCCTCGTC  
GCTCTCGTAGGTCTCGGGGACCGACTGCATGACCAGGTAGCGGTCCAGCA  
CCGAGGGTTGCTCTGGCGGGGGCGACTTGCCAGCTGCTTCTTCCGCGCG  
GCGACGATCGCGTCGATCATCGGCAGCGCCTCCTGACGAGCGGCCCTCGC  
CGCCGCCACGGCCGTCGCGTTCGTTGGTTCGGGTTGAGGAAGATCTCGTTGA  
10 ACAGCGCGTGGGTCCACGCCACCACCTTCTCGGTTCGGGATCTCGCCGACG  
CCGAGGTACCGGGCCATCGCGCCGCCCGGCACCCTGAGCGCGTAGTCACC  
GGTGAGATCGAACGGCTTGTGACGCCGACCTTGGCGAGCAGCCGGTTCTG  
CCTCGTCCACGACGATCTGACGGTAGCGGGCAGATCGGCGCGCGGGAAC  
25 GCGAGGCGCAGGAGCGACTTCTCGTGCTCGTACTTGGGCGAGTCGTTTAT  
CGCCAGGATGTTCTGGCCCACGTTCTCGACCAGCTTGGGCGCGATGTTGT  
CGACCGAGAAGACGTCGTTGGCGTTGAGGACCTCGACGACGTCGTTGTAC  
CGGGTCACGAGCGTGATGGCCGGGATGGAGAAGATGGGCTTCTCGCGCCG  
CAGCTGGCTGAGGAACGGGAGCGGCTCCTCCCTCAGCCACTTGAACACCA  
TGCCGGCCTCGATCTGCTTCCGCTTGACCGGATCGTTCTCGTGCGCCAGC  
20 GCGCTGTGGAGGGCCTGCAGGTAATCGAACGGCGCGCCTTGGCAGCGTC  
CGCTCGTCCCTCTTCTTCGATGTGAATGCTCATGGGGAGAATTCCTTTCT  
CGCATGCCGATCAGATCGCGACGCTCTGGGGGACCATCGACGGGAGCAGG  
TACAGGTACGGCTGCTCGCGGGCGCGGTTGCGCTCGGTGATCTCCCGCTC  
40 GATGTGCGCCAGGCGCTCGCGGTAGCGAGCAAACGCCTGCTTCGCCGCCG  
25 GATCTTGCAACAGGCACTTCATCGTCATCAGGCTCTCGCTGGTCAGCGGC  
GGGCCCATGCTGAGCGAGCGGCTGAGCACCATCTGGCGGATGCTCTGCGC  
GCGCCCCGGGAGGCGCTCCAGCGGCTTGAAGTCCGCTTCTCGCTTCCGT  
45 TCAGGACATCGCCGTAGGAGCGGTAGGTGCCAACTGAGCGTTTCGGGATC  
CAGGTGTAATAGTCCGTCTGCCCGAAGTTCACGGCGGCGTGATATGCGGT  
30 CGCCGTGAAGATGATGTTGGTGACGATCGCGATCAGGTCGTGAGGCTCG

5

207

10

5 TGAGCTTCTCGAGCTGATCGGCTCGCTCCGGCCGGAGGAGGCTATCCATG  
10 CCGCCGAGCTGGGGGGACACGAGCTCGTGGATCCACCGCTGCAGGCTGGC  
GTGCTCGACAGAGACCCCGGCGTCGGGTAGGCGATCTTCAGCACCTGTC  
CGACGTACTCCTGGATCGCGTCCCAGTGCAGCAGCGCTCGTCGCGGTAG  
15 TGATAGCCGACCAGGTCGCGGACGTGCGCGCCGACAGGTCGCGGGGGAG  
CGCGCTCTCGTAGAACCGCCACGGCTTGCCCGGTACCCCTTTGATGCCCT  
TGCCGGTGTAGGCGCGCTCAAGAGCTCGAACGAGCCCATGGTGGCCACC  
GAGCTCGTGATGTGAAGAAGCGCCCTCGCCCGAGGAAGCGCCGGCGAGC  
CAGCTCGTTGATGGCCAGGGTGTGAAGAAATGCGGCCTGAGCAGCTGGT  
20 10 GGAGCGGATGCGTCGCGGGCAGGTTGCGGTAGGTGCTCACCGCAACGGC  
TCCACGATCAGGTGCGCGTACAGCAGGTGGGTACCTGGCCCTGGTAGAT  
GGCGTCGGCGCTCGCGACGGCGATCTTCGCGGTGAGCCAGTCTGTCGACG  
25 GACCCGAAGGGGTGAAGATCTTGTGCGGATGCGCCCCCTTCCCGGGGCGC  
GAGTGACACCAGCCTGATGGCCACGGGCAAGAGCTCACCGGCCGCGGTCTG  
15 GTGCAGCATGCACGTGCGCGCCAGCGGGTACTTGCCAGCTCTTCTTGCA  
CGTCGGTGTGACGATGTCTTGAAGATGCGGTAGTCGAGGAAGTAGAGC  
30 TGCCCGCCCTCGCGCACCTCCTCCAGCGTGCAGCCGTGCGCGATCGCGAT  
CGGCTTGGGCTCGGCGCCGCTCACGAAATCGGCGAGATCGGCCGGGGTTCG  
CGCGCGGATGTGCGCCGGGTTGATCCCCACGAGGCGCTGCCGCCCGAAC  
20 TCGGCGTCCTCGGCCCAGCGCGTCGCCACGAGGGGCTTGCGGATGAAGGT  
35 CCACGGCTTGAAGAACTCCTCGAACTGATCGAAGCTCTCCAGTTGTGCA  
TGGACTCGAAGATGGCGCCAGCCGAGGTGCGACGTGGCCCTGAGGACG  
AACTTCCCCTCGCGATAGCGCTTGTACCCGTACTCGAACAGGTGAAGCGC  
40 CTGCGCGATTTGCAGTCCGGCAGTGTCTTCCACTTGCCGAGGTTGAGCG  
25 CCAAGATCTTCTTGATCGGAAAGCCTTCCCCCGGCGGTACGGAATCGCTG  
CCCGCCGGCAGGTTGGACCCGAAATTGCGCCAGTTCGGTGTGAGCTGGG  
CTCCCTCATGCTCGTGCTCGCTTCTCCGTCTCAGACGGACGGTGGATTGG  
45 GTGGTTACGTCAAACATCGCTCTCGCGTGCAGCGGTCCGAGCGCGCGC  
CGGAATGGTTCCGTCTCAGTCGCAACAGGACTCAGTACATCCAGCGCCGC  
30 CCCCCGTCTCGACCTGCCCCCGAGCCGATCGCGCCGCCCTTCATCGTG

50

55



5

10

15

20

25

5

10

15

30

35

40

25

45

30

50

55

GAATCGACAGGTGCGATTCCACGAAAAGCCGCCGCCGAGTTGCACGCG  
ACCGATGCTCACGCGTGCATTGTTGAGGCTGCTAGAAAACCGTGGAGCGT  
TCACGCATGTCAAGCCATTTTGTTCGGCGCCGCCGCGAGCGGCCGATGC  
CGCGCGCCCCCGCGCCGGCGTGTTCGCTCCCGACGTACCGCTACCTCGA  
CGACGTATGGCTTGAAGGGCAACCGCGCAAGTCGTCCGATTCTGTGCTCGT  
ATCCTGCTCCTTCCAGCAGGATTTCCCCGCCGCCAGCGGCACAAAGGTGC  
CAGGGCGAGCAGAAAGAGCGCTGCCCGCCCCCTCCCCGGCCGCGCTCGCCT  
CGATCAGCGCGCGCTCGTCGTGGTGCCTGAGACGTTTCGACGGACGTCAG  
TTAGTTAGCCTAGCTAACTTCAACACTGATGCGACTGATCGGGCCGACGC  
AACCAGCGCAACCGACGCAACCGACGCAACCGACGCGACCGACGCAATCG  
ACGCAACCGACGTGACGGACGCTGGCGACTCGAAGAAAACACGGACGCA  
CTCCACGTCATCGACGTCATCGACGTCATCGATGCGCTCGATGCAATCCA  
TGCACTTGACGCGATCGGTGCGAGCAGGCGACGAGGTCTCTCGTGAAAC  
ACCGAACCGAGTGCCGGTAGCGGGCGCGCCGAGTGTATGCTAGGCTCGG  
CCCTCTTGTCGAGCCGCGCGCTCGGCGGTGAGCGTGGGCTCGGGTGCC  
GCGGTATCCGGCTGAACCAAGGAGGAGCGAGCCATGCAGGCAGATGACGA  
CGCGACGATCTACAAGGTGGTGGTGAACCACGAGGAGCAATACTCCATCT  
GGCCGGCGGACCGAGAGAACCCGCTCGGCTGGACGGAGGCCGCAAGACG  
GGCAACAAGGCGGAGTGTCTGGCGTACATCCAGGAGGTCTGGACGGACAT  
GCGCCCCGCTCAGCCTCCGGAAGAAGATGGCCGAGAGCCCCTGAATCGCGG  
CCCGCCCGAGCGCCCGTTCGCGAGCGGCCGGCGGGCGGGCTCAGCCGTGTC  
ATCGTCGCGCTCGACCGGCCGCGTCCCGCGGATCGCGCGAGCCCGGCGG  
GGTCGTGCGCGCCGGCGCTTGTGCCGGGGCCCCGCTCTCGTACGCCTCC  
GTCATGCCGCCCCCTCGATCTGCACGTGCGCTTGTTCGGCGCCTCCGGCGC  
CGGCAAGACGGTCTCTTGGCAGCCTTCTACCGGGCGCAGACCCAGCCCT  
CGTTCCAGCAGGAGTACGCGTACAAGATCCAGGCGGTCAACAAGGCGCAG  
GGCAACCAGCTCTCTGGCCGGTTCTATCGCCTCGAAGAGGGCAGATTCCC  
GGACGGCAGCACGCGCTTCGACGAGTACGAGTTCGACTTCTTCCCGAGAG  
ATCTGCCCGAGCCGGCGGTCCGCATCCACTGGTACGACTACCCGGGACGC  
TGGTGGGAGGACGAGCCGGTTCGACGCGGACGAGCGGGAGGCGATGCGCCA

5

10

15

20

25

30

35

40

45

50

55

GGGCCTCATCCGGCTCGGGATGAGCCAGGTGGGCATCCTCCTCGCGGACG  
GCGCGAAGTACCGGGCCGAGGGCACCGGGTACATCCGGTGGCTGTTTCGAG  
CACTTCGCCGACGAGTGCAGACCGGCTGCGCCGGGCCAGCGCCGCCACGGG  
CGACGAGGTGAGCTTCCCGCGGGAGTGGATCCTCGCCCTCAGCAAGGCCG  
5 ATCTCTGCCCCCGGACTACAGCGCGCGGACTTCGAGCGCGAGGTCTGC  
CGGGACGCCGACGATCAGCTGGCGAAGCTCTGCTCGGTGCTCCGCGCCGA  
15 GCACGCGTTTCGGCCACCGCTTCATGCTGCTCTCGTCGGTCGCGCCCCCG  
CCGGCGCGCAGGTCGATCCGAGGACCTCGCTCGGCGTGCGCACCCCTCGCC  
CCCGCGATCCTGGTGAGCACGGTCGAGGGCGCGGTGCGCGAGGCGCAGGC  
20 10 GCGGAGAAAGGAGAAGTCGCGCGGASAGACGTTCTTCAGGGGCTGCCCG  
ATCTCGTGAGTTTCGTCGACTCCCTCGACGACTTCCTGCCGAAGCGATAC  
CAGATCGTGAGCAAGATCCTGCGGTTTCATCTCGATCAAGGACTTCGCGAC  
25 CACCCGGCTCGACCGGCTCAAGAAGATGCGCGAGGACGCGATCCGGAAGG  
GCGACACCTTCACGGCGGTCTTGACCGCGATGGTCGCGGCCCTGCGCGAC  
15 GACGAGGGCGCCCGCGCTACCACCAGAACCAGTGAGGTCGTCATGCCCG  
CGCCAGCGCCCCCTCGTCGAGACATCGCGCCTCCTCTGGAGGACGCGCGGC  
30 GAGCACTGGGATTACGAGTTTCATCTGTGTCCCGAGATCCCGGCGCTGCC  
CGCCTGGCTCTCGACGCTCGAGGCGATGCTCGCCGACGCCGACGCCGGCG  
CCGGGGAGCTCCGCTATGGCCTGCTCGAGATCGACGATCGCGGGCAGAGG  
20 GCGCCGCGCGCCTATCCCTACGTGSCCGTGAGGTTCTCTCGATCCGGCGCG  
35 GAGGGACTGGACCGGACGGCAGGTCCAGCACTTCGCGGCCTGGTTCCCGC  
CGGTCCCGCCCGAGGCGGTGCGGAGTTGCCAGAAGCGGTCCCCGCCGAC  
TGGCACCTTCGCGTGCTCGACGGCTCGCGGGACGTACGGCTCCGGCGA  
40 GGTGTTTCGGGCTCCCCGAGGCGACGATCCGCGCCTGGAAGCGGAGCCACG  
25 ACGAGAGCCGGGCGCGCGCGATGGCGATCGTCAAGGCGACGCCGCCG  
GTTTCGCTGGGCGGCGGCGAGGCGCGCCGTCGCGGTGGACGCGGGTGCC  
GACATTAAAAAAAAGCCGCCGA3CCGCCGGCCGCGGGGCTCCTCT  
45 CGGTGGGCGCGGTCCCTAGCGGCCAGGGCCGGCGATTGCGTGTTCGCG  
ATCGGCGCCATGATGCTCGCCGCTTCTGTGACTGATGCTCGCTTGCGG  
30 TGTGCGCCTCCTCGCGCCTGACG3CTGCGCCGCGCAGGCCATCCGACGG

5

10

15

20

25

30

35

40

45

50

55

GGGGTCGGCCCGGCCAGCGCCCGCCGGGCGACACCAGGGCATCGGCCCTC  
CGCTCGGGGCATCGATTGAGCTCTCCGAGCGGCGGTCCGTCTCAATCGC  
CGCAGAGCTCCCAACGGGCGGAGCAGCTCTGGCCGGTGACCGCATAGGGG  
TTCGTCTGGGAGGTCCACCACTCGCCCTGGAAAGGACGCGGGTTGCAGTG  
5 CGGGAGGCACTCCACCCACCCGACGAGCACGAGTTCCTTACCGAGACGG  
TCGGCTGAGCCGCGCAGAACCAGCGTTTTCTTGCGAACCAGCCGGATTG  
CACACATCGGGCGCCCCGCCGACGGGCGGATACACCGTGCGACCGCCTG  
GATGTCCACGGCGTCGAGGGTCTCGCTCCCGAAGCGACGCCGTCTGAT  
CCGAACACCCACTCCAGTAGGTTCATGATGGAGTCGTAGTCGTAGAAACCA  
10 GGGTTACGACGATGTACCGCCGGCTCGAGGGCCACCCGCTGGCGACGTC  
GCTCGCGGGGAGCGGCTCTCGTTGGCTGCAGGCGCTCGGGACCAACGGGT  
GATGCCACTCATGCATGAAGCCGATCGCATGACCCATCTCGTGATCGCG  
TACTGCTCCACGCAGTCGAAGCTGTATTGACCCGGGCTGTCTGCCAGTT  
GTACTTGATGCAACGGTTGAAGTCGGCGCCCCAGGGCTTGAAGTGGACCG  
15 AGCCGCCCTTGTTGTAAACGCCGATCGAGTCCGATTGGTTGGGCGCGTCG  
GGGTGGATCCTGACGCCGACGTAGGTTCATGCGAGTGGCCGGCAGGAGCGA  
ATCGCAGCTCTCCAGCCGGTGAAGCGAACCGAGCTCCAGCGTTCCACGC  
TGCCCTGGAGCGCGGTGCGCACGCGCGTGATGACGTCCGCGAGCGAGGGG  
TTGGGCGCATGGATCAGCCCGCCCGCGGCGCCGTGACCCCTCTGCTCCGC  
20 CGAGCTCGTGGGCTCGATGCAGACCGGGATCCGGACATGGCCGTACGCGT  
CCTCAGGCCAGCGACTCGCGCTGTGGAAGACGCTCGCCTCGGCGGACCGC  
GGCGCGGCAGAGCGGTGACGCGGGCGCCAGCGCTGCGAGGAGCAGCGG  
ACCGAGCGAAGAGCGAAACCGCACATGTCGTTTCAAGGCCCCGCGTCGTGC  
GGTGACCCGAGACAATCTCGAGCGGGCTCATGGACGCAAACGCGTTGCGA  
25 TGGCCTTGACGATGTTCTTCTCAATCGACGAGGGTTGTTCTGCTGAAC  
GCGGCTCCAGCGTGGAGCTCGACGCGGTTACCGGCTTACGCCGGGGCC  
GTGGACGAGACCCGAGCACGGGGGAGGTCCGAGCCGCACCGGCTCGCGGC  
GCCTCCACCCTGCACCTACGACGAGCCTGCCGCTCGGTTTCGCGGAAAAT  
GCCACCCCGCTGCCCAGCGGGCGAAGCGCGGACGAGGCGCTCGTCCCCAC  
30 GGTAGCGCCGGTGCCGCTGCATCCACCGCGCTCCTCCATGGGTCGCTGCC

5

10

15

20

25

30

35

40

45

50

55

CGCGGGTCGTCGAGGAGACGGACCCGGGGCGCGGATCCCTGGCTCGGCGT  
CGCATAGCTCGTAGGGCGGCCTTGAGCCGGCGGTACGAGCGGCGCAGTT  
CAGCAGCCGACCACGTGCACGCGGCGGCTCCGAGCTGCGCCGAGGAACC  
CTTCAAATATTAGATGGAATTCACAGGGTGGCTGAGAGACGGGGAGTAA  
5 GATCTCAGAGATCTCCCTGCCTACCCGCATCCCTGTTCAATTTCCGCCC  
ACAACGCGAACGGATGAGGAAATATCAGCCCGCGATCCCGACGGCCGACA  
15 GCATCAAAGGCCGCTCGAATCCAGGGGATTGAGCGGCCTCGGTGCGCGG  
GACCCCCGCGCGAGCCGCTTTGTCACTTTCACCACTTCAGAGCTTCG  
ATCATCTTCTCACCATAACGCGTGCCCATGATAACAACGGACGCATGATC  
20 GAAGTGGTACTGATCCATGACGTTGGTCCCCTGCTGGCTGACCCAGTACC  
10 CCATAGGCAGCATGTGCGCCGCTTGGTGACGAGGTTGTTATGACCCCGG  
CAGCACCCGCCCCGCGGGAGCTCTCCGAGAATGAAGGGAACGTCGTAGTC  
GACCCCCCAGGCTGCTTTCACCTCGTTATAGAGCTGAACGACCTTGCCGG  
25 GCCACGAGCTCTGGCCGTTGTCGGACTCACCTGGTGGAAGATGATGCC  
GCGAAGCGCGCTTCTCGGCCGTCTTCGCTTTGGCGATCTTGTTCAGAT  
15 CATCTGGTGATGCGAGCCACCAGTGATGAACGTGTTGATCGACTCGCCGC  
CATCTCAGCGGTAGCGACCAACCCGATCGTATCCCCCTCAGGCAGCTTTCCG  
30 AGCAGGGTCTTGCCGAACCAGATGCCCGGTCGACGGAGGTCGACAGGTT  
CCATCCTTTTTCACCAGGGCAATCGCTGAGCGGCGGATTGGCCAAGTTCC  
ACTGTCCGGCCGGCTGATTGCATCCGCGAGGACCTTGAGCCGCGCGTCA  
35 GAATTTTGTGCTGTCTGTTTGTCTGCGACACCAGCCATATTCGACTG  
GCCCATGAGCATGAAGATGTGAAACGTGCGACTCGCGCTCGGTGCGCCGC  
CGGTGCCTGCCCCGCTGCTAACGGATCCGGTCCCTCCCGTGGCGTCACCT  
40 CCAGTTCAGCGTTTCGTGCTGCCTGTGCGCTGCCCCCGGTCCCCGCGCT  
25 CGTGCTGCCTGCGGTGGCGCCGCGGTCCCCGCGCTCGTGCTGCCTGCCG  
TGGCGCCGCGGTCCCCGAGCCGCGCCCTCCGGTGTGTCGTCTCTACCG  
GTCGCGCCGGAATCGCCACAACCGGACGCGAGCGATGATGAAGAGGAATGG  
45 GAGGAGCAGGAACCTGGGTGTGCTCGGGTCGTGCGGTTTCATCTCGGTCA  
TGATCGTTACCTCGTCGCGCCGGGGCGGATCTGAAGAGCATGGCGGAAT  
30 CGGTAGGCCGCGTCGCGATGCCGGCGCGGCAACCTCGCCCGCAAAGAG

5

10

15

20

25

30

35

40

45

50

55

CTCAGCGCCGGGCTACCTTATCGCATCTTGGGCGCTTGGCGTCCAGGA  
TTCGGCCTTAGACAGCACAAAGCAGAAGACCTTTGACACTGGATTTTTTCA  
TCATCGGCGGCGCTCGTTCTTCGCTGCGCCTCAAGCGCCGACCGTTTCGTT  
TCGAAGCGAAGCGGTTTACGATCCGGGATCGCGAAATTTGAAACGGAC  
GCGTCGCGCGGGCAACGCAGGGGACTCATCACGAGGCAACCGCGCTGCGT  
CGCGAAATTGGCCAGCCTCTCGGAGTCCCTAGTTCCGTGCGTCAGACGCC  
TCACCCACCATGTCGAGCTCGGCGCGGCGCTCTACGTGCTTGAAAGACCT  
CCGCGAGCGCCGCGCTCTGCGCTCCGCGAGCGCACAGGCTCCCGTGGA  
TTCAGGGCAAGGCGGTCTGTGATCACGTCCTTCGTGGTCGCCGTGCTCTGG  
CCGGGCGGTCGACGCAGGACGGGTAGTACTGGCCGGCCCCGTTGAGCTT  
CTGGCAGACCCGGTCGCACGACCCGGCGATGCGGATCAGCCCGCACTCCA  
CGATCTGACCCCGGAGGTCACGTGGCCCGCGGCGCAGTCGCGCTGGTAG  
GCGCGCGAGTTGTGACCGTCGCGCTGTTGTAGCAAGCGTTGATATAGGG  
TTGCGCGGCGAACAGGTTGCCCCAGAAGGCCCCCTCGACGTCCGGATAAT  
CGATGAGCTCCTGGCTGGAGGAGAGCGTCTTCAGCGGATCCCGCAGCGAG  
CGGGCGGAGAGGAGCACCGGTACTTGATAGTAGTTCACGCGCGCCGCCAC  
GCAGCTGGACACGATGCGCTGCCCTGCGTCGTGAGCGGCCCCGCTCGCCC  
ACGCGGGCGCGACGCCGAGCAGCCCGGGGTAGCGCTCGTCGTGCTCTTG  
CCGTTTCGAGTCCGTCCACGAAAAATCGAAGGAGGCCGTGCTGCTCAGGGC  
GCAGCTCGCCGCGTAACGCAAGAAATCGCGCGCCAGCGCGCCGCTCGGCC  
CGGGATCCTGGATCGCGGCGAGGTTCCGCGCGCTGAGGCCGCTCAGGTTT  
AGGGCGTTTCAAGTTTCAAGGCGTTGAGGTTTCAAGGCGTTGAGGTTTCAAGG  
GTTTCGTGCTGAGCGGTTGCGGCCACGAGGGCCCCCTGGGATTCCCCCA  
CAGGCTCGCCCCACGCATCGGCGTCCACCACCTCGGCGGCGCAGCCCGAC  
AGCACCCTGCCCAACCAAGCACGATGAATGTCCGCTCGAGAGACATGGA  
TTCCCCCGTGTTCCTGGCGCATGACCCGACGGCGCCCTGCGCGCGGCGCG  
CGCGGGCTCCCATCGATTGCTGATGGGTTCAATATTCTACTTTTCCC  
GCGCTCTCGCGCCGGTGAAAGTCGCTTTCAGCGGCGGCGAGGTCGATGTCA  
GGAGCGTCCGACTCCGTGCTCTCGTCAGCTCCGCGTACCAGCGACGGAG  
TCGCCCCGCCATGACGGTCGGAATGGTAGAGGCGGCGGAGGGCGCGCT

5

213

10

CGAGCTGCGCCCGGGCGTCGGCGCGCCGGCCGCGGCGCAGGGCGGCGAGC  
GCCCCGCGCCTCGAGCACCTCGATCCGCTCCTGCCCCGACCGAGCAGCGCGC  
GGAGCGCTCCTCGAGCGCGGCCACGCGGCGCGGTTCGTTCGCGGGTCG  
CGAGCTCGATCATCGCGCAGAGCACGTCTCCGAGGGCTTCAGCGCCTCG  
5 CAGCCGGCGTTCGTCTCGCGCCGCGCGGAGCCGAGCGCGATCCGGCGAGC  
15 GCCCCGCTCGTCGCCC'GGTAGAGCGCAGGCGCGCGATCAGGAGCGTTA  
CGACGACCGGCGCGTGGCGATCGCCGCGAGCGGCGCTATCGCCTGGACG  
GCGCGCGCATGGGGCCTCGCGGCCCGAGATCGTCCATCAGGTACAGGTA  
CTCGGCGAGGTTGTAGCGGCCGACGAGCTCGAACGCGGGCTGGCCGAGCT  
20 10 CGCGCCCCGAGCGCGATGGTGCGCTCGAAATCGGCGATCATCCCGCGCGA  
TCGCCCTGGAGCGCCCCGCGGAGCCCGCGGTTGTTGAGCGCGGCGCCGAG  
GTGCATGAGATCGCTGCGCTCCTCGCAGCTGAGGATCACCGCGTCGAGGT  
25 CTCGCGCCGCTCCTCGACGCGGCGGAGGCTGGCCAGGATGAAGCCGAGC  
AGCAGCAGGGCGATGATGTGCGTCTCGTGGCCCTCGTCCCCGAGCCGCGC  
15 CGCCTGCGCCCGGCGCGCGTCAACACCGCGGCGCCTCGTCTCGCGGT  
CGGCGCGGTGGAGCGAGCGGCCACGCGGAGGAGCAGGCGGGCGCCGAGC  
30 AGGGGCGAGGCCACCCGCGCGGCGAGGCGCTCGGCGGCCGCGACCCGCTC  
GCGCGCGGCCCGGTACTCGCCCGTCCAGTCGAGGATCATGGCCTCGTCGA  
GGAGGAGCTCGATCTCGGCCCCCGCTCCGACGCGCGCCGCGCCGCTCG  
20 CGCGCCGCGGCGAGGTCGGCGAGGCGCTCGGTGTGGCGCCGAGCCGGAA  
35 GCGAGCGAGGCCCGCGCTCGGCGCTCCTCGGGGAGCAGCGCGCCGAGCA  
GCGCCTCGACGCGCCGTAAGCAGCCCTCGGCGTCGAGGTAGGCCGCGCGC  
GCGGCCGCGAGCTCGGCGCCGCGGCGAGGAGCGACGCGCGCGGGCGGT  
40 CAGGCCGCGCGCTCGCAGTGCGCCGCGAGCACGCGGATCGGCCTCGC  
25 CCGCGGCCTCGAGCCAGTCGGCGCGGAGGCGGTGGCCGAGCGCGCGATCG  
GCAGCTCGCCGCGTAACGCAAGAAATCGCGCGCCAGCGCGCCGCTCGGCC  
CGGATCCTGGATCGCGGCGAGGTTCCGCGCGCTGAGGCCGCTCAGGTTT  
45 AGGGCGTTTCAAGTTTCAAGGCGTTGAGGTTTCAAGGTTTCAAGGTTT  
GTTCTGCTGAGCGGTTGCCGCCACGAGGGCCCCCTGGGATTCCTCCCA  
30 CAGGCTCGCCCCACGCATCGGCGTCCACCACCTCGCGCGCGAGCCCGAC

50

55

5

10

15

20

25

30

35

40

45

50

55

AGCACCCCTGCCCAACCAAGCACGATGAATGTCCGCTCGAGAGACATGGA  
TTCCCCGTGTTCCCTGGCGCATGACCCGACGGCGCCCTGCGCGGGCGCG  
CGCGGGCTCCCATCGATTTCGCTGGATGGGTTCAATATTCTACTTTTTCCC  
GCGCTCTCGCGCCGGTAAAAGTCGCTTCAGCGGCGGCGAGGTCGATGTCA  
5 GGAGCGTCCGACTCCCTCGCTCTCGTCAGCTCCGCGTACCAGCGACGGAG  
TCGCCCCCCCATGACGGTCGGAATGGTAGAGGCGGCCGCGAGGGCGCGCT  
15 CGAGCTGCGCCCCGGGCGTCGGCGCGCCGCGCGCGCAGGGCGCGAGC  
GCCCCGCGCTCGAGCACCTCGATCCGCTCCTGCCCCGACCGAGCAGCGCGC  
GGAGCGCTCCTCGAGCGCGGCCACGCGGCGCGGTCTC3TCGCGGGTCG  
20 10 CGAGCTCGATCATCGCGCAGAGCACGTCCTCCGAGGGCTTCAGCGCCTCG  
CAGCCGGCGTCTGTCGCGCCGCGCGAGCCGAGCGCGATCCGGCGAGC  
GCCCCGCTCGTCGCCCTGGTAGAGGCGCAGGCGCGCGATCAGGAGCGTTA  
25 CGACGACCGGCGCGTGGCGATCGCCGAGCGCGGCGCTATCGCCTGGACG  
GCGCGCGCATGGGGCTCGCGGCCGCGAGATCGTCCATCAGGTACAGGTA  
15 CTCGGCGAGGTTGTAGCGGCCGACGAGCTCGAACGCGGGCTGGCCGAGCT  
CGCGCCCGAGCGCGATGGTGCGCTCGAAATCGGCGATCATCCGCGCGGA  
30 TCGCCCTGGAGCGCCCGCGCGAGCCCGGGTTGTTGAGCGCGGCGCCGAG  
GTGCATGAGATCGCTGCGCTCCTCGCAGCTGAGGATCACC GCGTCGAGGT  
CTCGCGCCGCTCCTCGACGCGGCCGAGGCTGGCCAGGATGAAGCCGAGC  
20 AGCAGCAGGGCGATGATGTGCGTCTCGTGGCCCTCGTCCCCGAGCCGCGC  
35 CGCTGCGCCGCGGCGCGCGTCAGCACCGCGGCGGCTCGTCCCTCGCGGT  
CGGCGCGGTGGAGCGAGCGGCCACGCCGAGGAGCAGGCGGGCGCCGAGC  
AGGGGCGAGGCCACCCGCGCGGCGAGGCGCTCGGCGGCGCGACCCGCTC  
40 GCGCGCGGCCCGGTACTCGCCCGTCCAGTCGAGGATCATGGCCTCGTCGA  
25 GGAGGAGCTCGATCTCGGCCCCCGCTCCGACGCCGCGCGCGCCTCG  
CGCGCCGCGGCGAGGTGGCGAGGGCCTCGGTGTGGCGCCGAGCCGAA  
GCGAGCGAGCCCCGCGCTCGGCGCTCCTCGGGGAGCAGCGCGCCGAGCA  
45 GCGCCTCGACGCGCCCGTAGCAGCCCTCGGCGTCGAGGTAGGCCGCGC  
GCGGCGCGAGCTCGGCGCCGCGGCGAGGAGCGACGCCGCGCGGGCGGT  
30 CAGGCCGCGCGCTCGCAGTGCGCCGCGAGCACCAGCGGATCGGCCTCC

5

10

15

20

10

25

15

30

35

40

25

45

50

55

CC3CGGCCTCGAGCCAGTCGGCGGCGAGGCGGTGGCCGAGCGCGCGATCG  
TCCTTGGTGAGCTGCGCGTAAGCGCCCTCGCGCAGGAGCGCCTGGCGGAA  
GGAGTACTCCTCCTCGCCGGGAAGCGGCCCTCGCGGTGGCGGACGCAGA  
GCTCCCCGGCGACGAGCGCGGAGAGGTGCTCCGCGAGCGGAGCGGCCTCG  
TCGCCCCGAGCAGGTGCGCGACGGCGCCTCGCCAGAACACCTCGCCGAG  
CACGCTGGCGGCCCGCAGGATCCGGCGCGCGGGGGCGCGAGCGCCTCCA  
GCCGGACCTGCACCATCGCCACCACCGTCTCGGGCAGCGCTCGCCGCGG  
CCCTCCGCCGTGCGCGGATCAGCTCCTCGAGGAAGAACGGCTGGCCCTC  
GGACTGGGTGACCAGACGATCGATGAGGGCCCCGTGCGCCGCGTCGCCCA  
GCGCCTCCCGCGCGAGCTGCGCGCACGCCCTCGGCGGGAGCTGCCTGAGC  
CAGAGCTCCTGCCGCCCCGCGCTCGGCCAGAGATCGGGGTACGCTTGC,

or their complementary strands,

(b) DNA-sequences which hybridise under stringent conditions to regions of DNA-sequences according to (a) encoding proteins or to fragments of said DNA-sequences,

(c) DNA-sequences which hybridise to the DNA-sequences according to (a) and (b) because of a degeneration of the genetic code,

(d) allele variations and mutants resulting by substitution, insertion or deletion of nucleotides or inversion of nucleotide segments of DNA-sequences according to (a) to (c), wherein the variations and mutants offer isofunctional expression products.

13. Peptide encoded by a DNA sequence according to claim 12 selected from the group consisting of



5

10

Seq ID No 92

>Contig11\_002 591 amino acids MW=63639 D pI=5.80 numambig=0  
MLDVWSTSDQVACRLHCAGAGPSASLELRYDASAGARRDAERLAERLAALLEDLSRHPER  
5 PVAQGEYVGPGERAEIEAWSRGPAMELPSACALHRWFEERAEQHPDVVAVRSEGKSLTYG  
15 ELERRANRLASCLRRRGVGLDTIVGVCVPRSEDMVVATLAVLKVGGAYLPDHEYPGERL  
AFMMRDARARLLVTHDAIADELPTGGWTTLLLDAAEAEIAACSDARPAVSPPPDSGAYVI  
YTSGSTGTPKGSLSHRAIVNQMWIORYWALTADDRVLLKAAFQFDVSVWEIFWPLSFG  
ARIVVARAGGHRDPEYLRRLVRDEGATTAYFVSSMLAAFLGGPEQPPASLRKVLVGGEA  
20 VPLDLVRRFYAKHGDGLINMYGPSEAAIAVTGCVLPSPRVTWVPLGAPVANAEEVFLDG  
10 AMRRPAIGALGDLYIAGAPLARGYVGQPLTAERFLPDPCARAAGGRMYRTGDVARFLPD  
GMLEFQGRSDHQIKLRGHRIELGDVEAQIRRVPGVGQAQAVVLREDAPGDARLVAYVVLGD  
25 DAAGDAPDVRAGLKASLSAYMIPSSVRLYALPMCSERLAFTGSSYAGCLL\*

25

15 Seq ID No 93

>Contig11\_007 361 amino acids MW=38862 D pI=10.42 numambig=0  
MSDHMTGFSLSPPQRAIRALDREAGAPGCRTLAVVAVTGPCDEGRLSAAALALAERHEI  
30 LRTRLVEAGRARPRRWSASRASRGRQQDDWVCSEAEQGERMSRLVARLSEDRGADDGLR  
VGLVRVGPEERRLVLAAPAWCVDEESIAPLVRELCASTAGAGAPPEQQYADVAEWLNGML  
20 ESEDAGDGRRFWAERRSHFGPPLHLAFSRGGAGAGAGSGRARVDLGQGGMAQVERWSSSW  
35 QVPQRIVLLALWASLLWRMSGGNEPEVTVAVRFDGRSLDALAGAVGPFARFLPVRIEISA  
SDTLADVARRLALAEAEAAAHQDAAPGVSHRMSWGLLRGGRAGAVARRRAGPRARRLEH  
V\*

40

25 Seq ID No 94

>Contig11\_012 882 amino acids MW=95015 D pI=12.69 numambig=0  
MARALYAQEAARSAFNACAAAARAGLDLGEVVFREGECDGRTLLGTALAQPALFAVEL  
45 ALARLWMSWGIEPAAMIGHSLGELVAACVAGVFTLEDAMSLVIDRGRFMQAAPAGSMLAV  
GLPAADVEGLLEAGLEIAAVNSPKLTVVAGPASAIRDLAARLEAREVFARPLQTSFAFHC  
30 ALIDGAVAPFLESVRRARLSPPEIPVVSNTGALLTDAEATDPAYWARHLRQPVRFSDGV

50

55

5

10

15

20

EALFASGHALFLEVGPGRCLTTLVRQTLACRGGAIASLGSTHAASEPASLAEALGQLWE  
AGHAVDWTARPRGRPPARSSACRRTRSSGRGTGSRRAAAPPPPSRRRPTRPSRPRRTPSR  
RRRSPARTPGPRSPRRRPSPSARSPAIWERPPRRPRDRRPRRLLRPARRLARRDPGDVPD  
PRAARRRAPAAGALPGADGRGARGPGRRRAPRRGAAARVPPDRADPPRRPAPAVVRAAPA  
5 VVRRSARAGQPGVQHPVRGARDGPARRRRAPPQPVRDRAAPRGAADDVQRQGRRAVPGRR  
ARGPGAVPDVRPRAPRGRGPRRGGLGARPRGEPRAVRPEPRAAPPRAGHPEAPRRRARHRA  
15 RRAPRRVRRLVGRRLRRRARRALRRLRAGPAFPAAGAPGAVRRLRRSPAGVAVRRGPGR  
APVLDDREALGRAAPGPRAGRPRAGGAKDVARGAAEPRRRGRAHPPDQGVLRAGGDLPLHG  
PARRVQARPAPAHRRARGPRRGHRRREPEPRRDADDRLLRQPARAPDRLRGRPDVRRAGP  
20 PRARRRRARGVRAPGPAVRPARRGPAAEGRRRAPLRREVRHAQARPPDEARGARARGA  
10 GGRGDHDGVRLRPHGRRGRRLVPLRRRAQRAVFGDGRQFP\*

25

30

35

40

45

50

55

Seq ID No 95  
>Contig11\_021 1213 amino acids MW=131017 D pI=12.40 numambig=0  
15 MRGRRRRRAAPHLRGARPARRGDRRGGLALRSDREARRRRGHGEGLGAGRRARRAARRRR  
LPAPGPSPPRGAPQAPPGARRGPPRPHAVGGRDDRLARGHRAPRRRRRGRAVARAARRAP  
30 AARRKHGRPGLRDLHVRVHGAAQGRDDRPPRRGEHGPRHQPLRRRPPGGPGARALVAELR  
PVGLRRVRDARRRRRRRDPRPYRLGSGALARARGARAGDRVELGPGADGDAHGRVPRRG  
RPGAVVAPPRHDERRLDPAEAPRSHPRGLPRAPRREPRRRDRGVLDVDRPPDRGRRPVGA  
20 QHPLRPPAREPAHLRARRGAGAVSDRGPRRDPHRRDRRRARILARRGADPGAVPEAPHDR  
35 RAAVQDRRPGPLRRRDHRAARAHRPSGEDPRLPHRARRDRGRPRAAPLGRAGGRGGEDR  
SVGREAPGRVRRRRRRRRRAPRLRPEEAAGVHDPRGGRPPGAAAERERQGGPRRPAGS  
RRGRAARRGRRPADRDRAAHLRGARGGLAGRGRRRDROPLRARLHLAAPRARAAPARRAH  
40 RGARARRGRRGAGRVADRPVVPDHRATGAAARRGDGQGRAGRRRGAARRGAPRRATPPG  
25 ARIAMSEPIETEDGGSIDIAIVGMAGRFPAGPSVDALWENVRRGVESIARFPESEREPPV  
GASAAPGAPVVCAGGLDDIDRFDASYFGYSPREAQLMDPQQRLFLECAVAALDAGCDP  
ARFPGAIGVFGGCGSNTYLLQLLSHPDLAATVDPHALMLASEKDYLATRVSYKLDLHGPS  
45 VVVQTACSTSLVAVHMACESLLGGQCDLALAGCVSIGIPQKRGYPYVPGSICSPDGRCRP  
FDARAEGTVGGSGVGIVALKRLAJALDRNTVHAVIRGSANNDGGRKVGFMAPSVDGQA  
30 AAISEAQSVAGVDPGSGIGYVEAHGTATAIGDPPIEVEALTQAFRRKTPRKAYCALGSIKAN

5

10

IGHLDAAAGVAGLIKAAHVVRSGEIPPCVHFEAPNPKLDLAASFFVPREAAPWPRELRP  
RRAGVSSFGIGGTNAHVVEEPPPLPPRAPAPERDHLVTLSTARTPEALSTACAQLAAHLE  
ATDVPLDDVAFTLTQTGRAEHYPYRAVVARTRAEAIQGLAREGASALARPDEPRPSSRSRA  
RARRPSGWPASTRRRRRSGAPSTRARRRRGRAASISARSSSARARATGARCSAPRWRSP  
5 RSSPSSSRSPGSG\*

15

Seq ID No 96

&gt;Contig11\_026 3079 amino acids MW=332984 D pI=5.97 numambig=0

20

10 AEYWRKALDGATTAIDLPRDRARHDAGARRGRAHAITLPKPLTGALARLARERGTTLFSV  
LLSALTVLLHRASGQSDLVVGVPASAGRHDDESARAFGYFVQMLPVRVALRGAASFDALVA  
RVRDAFLDALAHGDSALRHLLARAQGEAQRDALFDVAFQSTPPSLDARLSALAIGVGD  
25 VRIAQGELELTTLADEQAAAEFDLALFAAELDAGIALRFEYDQQLFDPATIERMARHFVV  
LLESAVEHPGRPLSELRLMSDAERALLDDWSGAAAARQAASAPAPACVHALFEAHAARQ  
15 PDATALEFGHQRTYAQLSTWSTELALWLRDRGVGPGSVVGVCIERSPRMVAAQLAVLKA  
GAAYASLDPANPPARLAEMLADCRASALTSSQASHKLTAAPCPVHLVQDGACAPSTHIP  
30 LVS RPDDLAYVLFTSGSTGTPKGVCVRHASLSRLVSFLHLRLDLSPSDRWTQVASSGFDA  
SVYEIWTPLACGAALLLADDDALRSPTALVSWLVAQRATLSFMPTPLAEACFEQDWTGSA  
LRAMTVGGDKLHPLRRPLPFRLFNMYGPTEATVITVAEIADLGAEPPLGRPVDALVYV  
20 LDPHMQPVPPGALGELYIGGACLAQGYTRTDLTAEERFLPDPFGQPGARLYRTGDLVRWRP  
35 DGQLAFAGRDEQVKLRGRVVELGEVESVLRRLPGVREGIVVLHCQGSAAHLIAHVVPDA  
HPPSERDLREGMARLVDPDALVPAHFVLLPALPMSLSGKVDKLLPAPPAHADYEPPSGE  
LELELAHIWQSVLHLDRVGRHDSFFDLGGHSLAMQVLGRIESSLGIRTTLRTLFEHPTL  
40 AQLASHLSSGAASTSAAAATALERGLTRPDGPSSSRVATPEEPFALTEGQRAMWLECQKS  
25 ADGALYNLGRTVRLGAGVDVAALRRAFEGGLVERHEALRTTFLTRDGHPLQQVHRHVALEW  
AEEPAMALDEREIVARADEVRRRAFDLERGPLLRVHVWRRGEGQPPLLTVVVHHLVVVDYW  
SFALLVRELGELYSALRAGRPPQLPPSSFFAAGVSCPSPREAAGGAEYWRKALDGATTA  
45 IDLPRDRARHDASPRRGRAHAITLPKPLTGALARLARERGTTLFSVLLSALTVLLHRASG  
QNDLVVGVPASAGRNDDESTRAFGYFVQMLPVRVALRGAASFDALVARVRDAFLDGLAHGD  
30 SALQHLLAEPRGAARRGGALFDVAFQFQALPSLDPRLAALTGAEDVRIAQGELELTTL

50

55

5

10

15

20

25

15

30

35

40

25

45

30

50

55

ADEQAAAEFDLALFAAELD SGIALRFEYDQQLFDPATIERMARHFVLLLESAVEHPGRPL  
SELRMLSDAERALLLDDWSGAAAAARQAASAPAPACVHALFEAHAARQPDATALEFGHQRF  
TYAELSTWSTELALWLRDRGVGPGSVVGVCIERSPRMVAAQLAVLKAGAAAYASLDPANPP  
ARLAEMLADCRAALVLTSSQASHKLTAAPCPVHLVQDGACAPSTHIPLVSRPDDLAYVLF  
TSGSSGTPKGVCVRHASLSRLVSFFQHLLALSPDRWTQLASSGFDASVYEIWTPLACGA  
ALLLADDDALRSPTALVSWLVAQRATLSFMPTPLAEACFEQDWTGIALRAMTVGGDKLHP  
LRRPLPFRLFNMYGPTEATVITTTVAEVDLGDPEPLGRPIDSALVYVLDPHMQPVPPGVL  
GELYIGGACLAQGYTRDTLTAERFLPDPPGQPGARLYRTGDLVRWRPDGQLAFAGRRDEQ  
VKLRGRRVELGEVESALRRLPVAVREGVVVLHGQGSAAARLIAYVVPADPPSERDLREGMA  
RLVPDALVPAHFVLLPALPMSLSGKVKLLPAPPAAHADYEPSPGELERELAHIWQSVL  
HLDRVGRHDSFFDLGGHSLLAMQVLGRIESSLGIRTTLRTLFHPTLHQLADRLSSGAAS  
TTAAATVPASEIAPSLGRAPADEPYPLSYEQERLWVLEQLLPGGTAYNVVQAVRLRNLV  
DVDALSSALAAALVRRHWSLRTVTFVASPTPAQKICEPEAAPAEVVDLRGTPDEAEAAARA  
WASREQATGFDLARGPVFRARLFRLDHDCVVLVLSHHIVTDAWSFQPLVRDLAELYRRA  
RGGGPADMPPELPLQYVDFAVWQRRHLAGKRLADKLAHWATLRLGLPVLELQTDPRPPVQ  
TFRGAERVLPLDARLVAQLDELARSRGATRFMVLLAALGVLLRRSSGQDDLAIGTAVANR  
PRPELEPLVGFFVNTIVMRDLGGDPTFEELLSRARKVALEAFEHQDAPFEKVVEAVNPR  
RDLRSRPLFQVMLVVQNAPTEALELGEVRIEPLDLPVEATRFDLRFVSVEPRGGRDVISLQ  
YNVDLFDAAATIDRMLATMQSVLSRATQDPAQVRALSVAPEDRERLVAWNDTAVATPDH  
LRLEEPFFFERAVEQPDACAVVDAERRLTYGELARRAEIAAAAASRSGATANALVAVVMEK  
GWEQVAAVLVGLRAGAAYLPLDPRLPEERLRHLLHAEVRLVLTQSAVDGTIAWPAGIER  
LAVDADERWREQPVARRPPGGSTDDLAYVIYTSGSTGLPKGVMIDHRGAVNTVLDINRRF  
DVGPEDRVLA LSSLSFDLSVYDVFGLAAGGAVVIPDRTRASDPGHWR ELVERERVTVWN  
SVPALMEMLMDASPGAGDPALSSSLRVMMSGDWIPLKLPDRIRAACRAPRVVSLGGATEA  
SIWSIAHPIADVDPAWRSIPYGRPLANQHTYVLDEGLEPCPIGVPGEIHIGGIGVALGYW  
RDEARTRERFLKHPTTGERLYRTGDLGRYFADGTIELLGRTDHQVKIRGFRIELGEIEAA  
LAQHPSVEQAVVAAKTDPSGEKRLVAYVVGADGDGAALRDFVRKKLPEYMI PAEVVVLPA  
LPLSANGKVDRAALPDPAAVAPRAAAVAPRTATERLIASVLAEVLQVEAVGVTDNLFELG  
FTSLLLVRQRLLAERIAARAPDEGAAAQAVSLTDLFYPTIEQLAQRLDAATVKAEPAD  
VGAQRAEARRDARRRRGRG\*

5

10

Seq ID No 97

>Contig11\_011 544 amino acids MW=60164 D pI=9.10 numambig=0  
MMSRIRAQLGVLEPLRALFQGPTVAALAAQVDAARRGEARRREFPPIARIPRDGPLPLSF  
5 AQHRLWFVDQLEPGSPAYNIPFVVRATGRLDVDALRRSLFEIARRHEALRTTFSARDGVP  
15 FPVVAPEARVPFRMSDLEHLAGEALDAAVSALVLEESLAPFDLSRGPLLRVRVIRKRHDE  
HVIALVVHHVFDVWSVGVFVGELAAALYGGFAQQQPSRLPELPAQYVDFAAAQRAWLSGE  
VLEGELRYWTTKLSCALRRARVPVDHEPAGRRTWRGARRSLDAGAELTRQIKAFCEAREAI  
SPFMALLAAYKLVLHQRTGLEDLVVGTDVANRRNVETEPMIGFFVNQLVLRDCCGGDPTF  
20 GALVRRVRDVALEAFEHQDLPFDRLEALRPKGAVGHVPLFDAKFVMRNVHVPMPKLEGL  
10 ELEALEGEATTTAFDFVLTVAEAGGSFRFGVEHSSSELYRAATVDNFLSDYRQILATATAR  
PDTFVSELRGELERAAAARELERKAARGAALDKLTSARRRAVTLPRPGAPGEAKTSPKD  
25 DLDE\*

25

15 Seq ID No 98

>Contig12\_001 514 amino acids MW=56145 D pI=8.82 numambig=0  
PPAVRRYVADRRPEQLPALAPEEREAAARRLSALGAAPPQVRRRGLTRAPLSYQGSRIYF  
30 LEQLSPGKPLFNVPGAVRLRGPVDVARLSAAFGEIVRRHDALRTSIANVDGELLQIAQPH  
AGFALDVVTSTPEEAAELDRRLRAEAWRPFAIGAPPLLRATLFRLAEDEHVLLVTMHHV  
20 SDDWSLGVILRELLALYAGRSLPFPRLQVSDFAAWQREMVESGALDGQRAYWRERLRGLS  
35 RASISAGGGAEAPSHDPSGAIEEIALSPDKAAALEALARREGATLFMVLLALLDLVIHAR  
SGALDIAVGTPIANRRRPELEDVVGLLTNTLVIRVDLARAGAFRDVLARARVQALDAFAN  
QDIPFDVVTQDLKQERDHAQHPLFRVWLALQNAPKPALEVRGLRVEPLPLRPELVHFEVA  
40 LLLWPADDGSSVVGHFEFRRDRVDEGARKEIAAAFTHLVDAVIARPDAPVSTLVEGARAEA  
25 ARAQAALGEAFARAATARLGQLRRRSAGDRTPRE\*

40

Seq ID No 99

>Contig12\_009 582 amino acids MW=65555 D pI=8.72 numambig=0  
MREPSSTPNWRNFGSNLPAGSDSVPPGEGFPIKKILALNLGKWKDTAGLQIAQALHLFEY  
30 GYKRYREGKFVLRATSDLGLGAIFESIDNWESFDQFEEFFKPWTFIRKPLVATRWAEDAE

50

55

5

10

15

20

25

30

35

40

45

50

55

FGRQRLVGINPAHIRRATPADLADFVSGAEPKPIAIAADGRTLEEVREGGQLYFLDYRIFK  
DIVD TDVQEELGKYPLAPTCMLHQTAAGELLPVAIRLVHSRPGKGAHPDKIFTSPSGPSDD  
WLTAKIAVASADAIYQGQVTHLLYAHLIVEPFAVSTYRNLPATHPLHQLLRPHFFNTLAI  
NELARRRFLGRGRFFDITSSVATMGSFELLTRAYTGKGIKGYGGKPWRFYESALPRDLA  
RDVRDLVGYHYRDDALLHWDIAIQEYVGQVLKIAYP TPGSLSSDASLQRWIHELVSQ LGG  
MDSLLPPERADQLEKLTSLDDLIAIVTNIIFTATAYHAAVNFGQTDYYTWIPNAQFATYR  
SYGDLNGSEKRQFKPLERLPGRAQSIRQMVLSRSLSMGPPLTSESLMTMKCLLQDPAAK  
QAFARYRERLAHIEREITERNRAREQPYLYLLPSMVPOSVAI\*

10 SEQ ID No 100 (>ORF1)

VSQRTSCYLRGGGVCSMNDAFLALERNERNRPSTVIDLLRQRAEAPARPIYCFLESGDVEAG  
ATWVTLREIDERARTVAALLQASGVAPGARALLYPPIEYITAFFGCLYAGVRTVPAYPPDL  
GRLERTLPRVASIVADARAEEAALTSSAVAGIVASLPASAAAAALQRLRWIATDGPSPGPIEGP  
GAALRPESVAFLOQTSGSTGEPKGVMLTHGNLLHNSRLIAHGFDLTSPDPVGVWLPPYHDMG  
LIGGILQALYRRIRVALMSPLSFLQRP MRWLRAVSALGASVSGGPNFAYDLCVRKSSEERAA  
LDLRSWEVAF TGAEPVRADTLDRFARAFVSGFRREAFYPCYGLAEATLIVSGGARAEAPVLA  
RLAPEEVELGRAVASAAEGARVFGSGRALDPRAVAIVDPAGNELGPGEIGEIIWVSGPSVAVG  
YWGRPEETEATFGATLAGSAAPRYLRTGDLGLRGGELFVVGRSKDLIILGRNHFPQDIEKT  
VESSHRAVRPGCSAAFSVEHEGEERLAVVCEVDPRVAADPREIVAAREAVTAEHQLVAHAVAL  
IAPGALPKTSSGKVRRECRRAFLEDALGERHVAFAPELLDDASPPDDAPPETEESGRSLLD  
ALRSTLARALRLDAGQIDDALPISRFGLD SLAAVELQHAFQVRTGRAIPLTSILRGGSRLRLTR  
EITRLDGPSSPRVATPGGAVCADRWGTGRFGSSAISRPMERFTTWAGRSGSVPAFKRVDLRR  
F

25 SEQ ID No 101 (>ORF2)

VYSSAYVLFVACAGTTRVASAPETAGFPLECVGDDGTVLGPDSFVVGYTQVYVFKKERLNTNP  
PIDGFTLKLDGNEVAPGEDGLPVVKRCVRSEEQAQCGGRTEPAEDECTTYEIEAVVPEKAAEV  
DEEAAGLGGPPAREAIWVDYYTDGGEFDGARRLVSDTTGASRGGNGTTWTPPSEPGRVSLWAV  
VHDTRGGASVTRREVQVE

30

5

222

SEQ ID No 102 (&gt;ORF3)

10

VVGTVLSAGTGEPLPDIAVTLVRPDGGREEAKTDQGGKFRFKNLPPGKYRVEVAAAGFEPFAA  
EEEIAAGEAIEVRYRISLAAPQDGKAPGIEVTVQGERPPREVTRRTIERREIDRIPGTGGDAL  
RSLQSLPGVARSGFGLLIVRGSAPODTLTFVDRTVPVPIIYHFGGLSSVVPTEMLEKIDFYPGN

5

FSAVYGRAMGGIVDVGLRSPKQDGKYHGVVQLDLIDGRVLLGEPVPFLKDWTFIAAGRRSWVD

15

AWLGPVLKEAGSSVTQAPVYYDYQFVLEGRPSASERVRASFYGSDDAFKITLDKPPPEDEPALT  
GDFGLHTAFQRFQLSYENRIGSRDRLLWSMALGRDIADFEISPLAFNVVSTSLDLRLELSHRF  
ARYLTMNVGTDLSGGVATVNIRAPSQQPAGHPSNQPFSTYPFQDRSFDGAYSRAAYAELEVV

20

10

PSPRARIVPGVRVDYALDTQTLDVSPRVNARYDIRSGFPRTTAKGGVGLYYQAPQFAESIEPF  
GNAELKSNRAVHYGLGVEQEITPQIEVTLDDGFYKQLDRLVVFSPKDDYADGTGYAVGGELLL  
KYKPDERFFGWAAYTLRSVRKDGPDDEEHLTQFDQTHVLTVLGSLRLGRGWELARFRLVSGN  
LQTPYVCDPEEKGCNPNRVNAIYHASSARYSPIPLGGDYSERMPLFHQLDIRADKTWKFKRWQ  
LGLYLDIQNVYNMAAEGISYNFNYTKREYVTGLPFLPTLGLRGDF

25

15 SEQ ID No 103 (&gt;ORF4)

VIAVDNNEAVDAVKDKTSAAFVGDATVHKVLEGIGAQYVETAIVTFGEHFEPVSLCVASLVR  
MGVRIIARAATDRQADILRAVGATRVIQLETEMGRRVGADITMPLAQDLLDLASHYRVVPWNA  
HGPLVGQTLGASKIRQRYRINVLGVRPHTNKRPGDKPRLEAPTPDYVIRDGDTLLLVGDSDDV  
SRFVAEVGG

30

20

SEQ ID No 104 (&gt;ORF5)

SGSSGGGSSAEGSRCQPSGGGPHWLEGETVTFPVTCASGLALAGDAFEVGPPLPEGAAYDPIA  
REVTFSPLDQAAVYDIEIRVAQTSEVGRVKGVADAFADPSNPVVDPTRYPEEYGLPVLFL  
SPVPEDKEYAPATVIYRGHTYAAEAEELRGESSLSYPKRSYTLKFPKDDKFNEPDEAGGFTDRR  
KVVLIITTFDDNSYVRQRLAYDLWNRDLPEHIQIKTYSAVLYLDGEYAGLYTVADHVDGYLMED  
HGYPQDGNLYKAVSHDANFALTDRSGDPKDTLHDGFEKKEGAPAEGEPEAFSDLEDLVSFVAE  
SDDATFAAEIGSRIDLRDYEDWWIFVTFIVANDSAGKNSYHYRDPADGVFRYAPWDFNASFG  
QSWETEREPASDRVDYRDVNLFLERLLEEPSIGDPLRARYDQVLRGALAEAEIHAIVDGYVER  
IDASARRDEARWGEAYRSYEGWSWRDDFTTYEEEIAYLK

35

40

25

45

30 AWISERWQHODELY

50

55

5

10

SEQ ID No 105 (Contig 11 &gt;ORF1)

VLDVWSTSDQVACRLHCAGAGPSASLELRYDASAGARRDAERLAERLAALLEDLSRHPERPVA  
QGEVGPGERAEIEAWSRGPA MELPSACALHRWFEERAEQHPDVAVRSEGKSLTYGELERRAN  
5 RLASCLRRRGVGLDTIVGVCVPRSEDMVVATLAVLKVGAYLPDHEYPGERLAFMMRDARAR  
15 LLVTHDAIADELPTGGWTTLLDAAEAAEIAACSDARPAVSPPPDSGAYVIYTSGSTGTPKGSL  
ISHRAIVNQMWIQRYWALTADDRVLLKAAFQFDVSVWEIFWPLSFGARTVVARAGGHRDPEY  
LRLVRDEGATTAYFVSSMLAAFLGGPEQFPASLRKVLVGGEAVPLDLVRRFYAKHDGDLIN  
MYGPSEAAIAVTGCVLPSPDRVTWVPLGAPVANAEVFVLDGAMRRPAIGALGDLYIAGAPLAR  
20 GYVGQPLTAERFLPDPCARAAGGRMYRTGDVARFLPDGMLEFQGRSDHQIKLRGHRIELGDV  
10 EAQIRRVPGVGQA AVLREDAPGDARLVAYVVLDGDAAGDAPDVRAGLKASLSAYMIPSSVVR  
LYALPMC SERLAFTGSSYAGCLL

25

SEQ ID No 106 (Contig 11 &gt;ORF2)

15 MSDHEMTGFSLSPPQRAIRALDREAGAPGCRTLAVVAVTGPCDEGRLSAAALALAERHEILRT  
RLVEGRARPRRWSASRASRGRQQDDWVCSEAEQGERMSRLVARLSEDRGADDGLRVGLVRVG  
30 PEERRLVLAAPAWCVDEESIAPLVRELCASTAGAGAPPEQQYADVAEWLNGMLES EDAGDGRR  
FWAERRSHFGPPLHLAFSRGGAGAGAGSGRARVDLQGGMAQVERWSSSWQVPQRIVLLALWAS  
LLWRMSGGNEPEVTVA VRFDGRSLDALAGAVGPFARFLPVRIEISASDTLADVARRLALAEAE  
20 AAAHQDAAPGVSHRMSWGLLRGGGRAGAVARRRAGPRARRLEHV

35

SEQ ID No 107 (Contig 11 &gt;ORF3)

MSRIRAQLGVELPLRALFQGPTVAALAAQVDAARRGEARRREFPPIARI PRDGPLPLSFAQHR  
40 LWFVDQLEPGSPAYNIPFVV RATGRLDVDALRRSLFEIARRHEALRTTFSARDGVPPFVVAPE  
25 ARVPFRMSDLEHLAGEALDAAVSALVLEESLAPFDLSRGPLLVRVIRKRHDEHVIALVVHHV  
VFDVWSVG VFGELAALYGGFAQQQPSRLPELPAQYVDFAAAQRAWLSGEVLEGELE RYWTTKL  
SGALRRARVPVDHEPAGRRTWRGARRSLDAGAE LTRQIKAFCEAEISPFMALLAAYKLV LHQ  
45 RTGLEDLVGTDVANRRVETEMIGFFVNQLVLR TDCGGDPTFGALVRRVRDVALEAFEHQD  
LPFDRLVEALRPKGAVGHVPLFDAKFVMRNVHVP MKLEGLELEALEGEATTTAFDFVLTVAE

50

55



AGGSFRFGVEHSSELYRAATVDNFLSDYRQILATATARPDTVPVSELRGELERAAAAARRELERK  
AARGAALDKLTSARRRAVTLPRPGAPGEAKTSPKDDLDE

SEQ ID No 108 (Contig 11 >ORF5)

MSEPIETEDGGSDIAIVGMAGRFPGAPSVDALWENVRRGVESIAARFPESEREPEPPVGASAAPG  
APVVCAGLLDDIDRFDASYFGYS PREAQLMDPQQLFLECAVALEDAGCDPARFPGAIGVF  
GGCGSNTYLLQLLSHPDLAATVDPHALMLASEKDYLATRVSYKLDLHGPPSVVVQTACSTSLVA  
VHMACESLLGGQCDLALAGGVSIGIPQKRGYPYVPGSICSPDGRCRPFDAEAGTVGGSGVGI  
VALKRLADALDRNTVHAVIRGSVNNDEGGRKVGFMAPSVDGQAAAISEAQSVAGVDPGSGIGY  
VEAHGTATAIGDPIEVEALTQAFRRKTPRKAYCALGSIKANIGHLDAAAGVAGLIKAAHVRS  
GEIPPCVHFEAPNPKDLAASPFVPREAAPWPRELPRRAGVSSFGIGGTNAHVVLEEPPPL  
PPRAPAPERDHLVTLSTARTPEALSTACAQLAAHLEATDVPLDDVAFTLQTGRAEHYPYRAVVA  
RTRAEAIQGLAREGASALARPDEPRSSRSRARARRPSGWPARSTRRRRSGAPSTRARRRRGR  
AASISARSSSARARATGARCSAPRWRSPRSSPSSSRSPGSG

SEQ ID No 109 (Contig 11 >ORF6)

VVDHHVVVEYWSFALIVRELGELYSALRAGRPPQLPPSSFFAAGVSCPSPREAAGGAEYWRK  
ALDGTTAIDLPRDRARHDAGARRGRAHAILPKPLTGALARLARERGTTLFSVLLSALTIVLLH  
RASGQSDLVVGVPSSAGRHDDESARAFGYFVQMLPVRVALRGAASFDALVARVRDAFLDALAHG  
DSALRHLLARAQGEAQRDALFDVAFAPQSTPPSLDARSALAIGVGDVRIAQGELELTTLADEQ  
AAAEFDLALFAAELDAGIALRFEYDQQLFDPATIERMARHFVVLLESASAVEHPGRPLSELRLMS  
DAERALLLDWSGAAAAARQAASAPAPACVHALFEAHAARQPDATALEFGHQRTYAQLSTWST  
ELALWLRDRGVGPGSVVGVCIERSPRMVAAQLAVLKAGAAAYASLDPANPPARLAEMLADCRAS  
LALTSSQASHKLTAAPCPVHLVQDGACAPSTHILVSRPDDLAYVLFTSGSTGTPKGVCVRHA  
SLSRLVSFLHLRLDLSPSDRWTQVASSGFDASVYEIWTPLACGAALLLADDDALRSPTALVSW  
LVAQRATLSFMPTPLAEACFEQDWTGSALRAMTVGGDKLHPLRRPPFRLFNMYGPTEATVITT  
VAEIADLGAEPPLGRPVDSALVYVLDPHMQPVPPGALGELYIGGACLAQGYTRTDLTAERFLP  
DPFGQPGARLYRTGDLVRWRPDGQLAFAGRRDEQVKLRGRRVELGEVESVLRRLPGVREGIVV  
LHGQGSAAHLIAHVVPDAHPPSERDLREGMARLVPDALVPAHFVLLPALPMSLSGKVDKLLP  
APPAAHADYEPPSGELELELAHINQSVLHLDRVGRHDSFFDLGGHSLAMQVLGRIESSLGIR

5

10

15

20

25

30

35

40

45

50

55

TTLRTLFEHPTLAQLASHLSSGAASTSAAAATALERGLTRPDGPSSPRVATPEEPFALTEGQR  
AMWLECQKSADG  
ALYNLGRTVRLGAGVDVAALRRAFEGIVERHEALRTTFLTRDGHPLQQVHRHVALEWAEPPAM  
ALDEREIVARADEVRRRAFDLERGPLLRVHVWRRGEGQPPLLTVVVHHLVVDYWSFALLVREL  
5 GELYSALRAGRPPQLPPSSFFAAGVSCPSPREAAGGAEYWRKALDGATTIDLPDRARHDA  
SPRRGRAHAITLPKPLTGALARLARERGTTLFSVLLSALTVLLHRASGQNDLVVGVPSAGRND  
DESTRAFQYFVQMLPVRVALRGAASFDALVARVRDAFLDGLAHGDSALQHLLAEPRGAARRGG  
ALFDVAFQFQALPSLDPRLAALTGAEDVRIAQGELELTTLADEQAAAFFDLALFAAEELDSG  
IALRFEYDQQLFDPATIERMARHFVLLLESABEHVGRPLSELRLMLSDAERALLLDDWSGAAAA  
10 RQAASAPACVHALFEAHAARQPDATALEFGHQRTYAELSTWSTELALWLRDRGVGPGSVV  
GVCIERSPRMVAAQLAVLKAGAAAYASLDPANPPARLAEMLADCRAALVLTSSQASHKLTAAPC  
PVHLVQDGACAPSTHIFLVSRPDDLAYVLFTSGSSGTPKGVCVRHASLSRLVSFFQHLLALSP  
RDRWTQLASSGFDASVYEIWTPLACGAALLLADDDALRSPTALVSWLVAQRATLSFMPTPLAE  
25 ACFEQDWTGIALRAMTVGGDKLHPLRRPLPFRLFNMYGPTEATVITTTVAEVADLGDEPPLGRP  
IDSALVYVLDPHMQPVPVPGVLGELYIGGACLAQGYTRDTLTAERFLPDPPGQPGARLYRTGDL  
VRWRPDGQLAFAGRRDEQVKLRGRRELGEVESALRRLPAVREGVVVLHGQGSAAARLIAYVVP  
GADPPSERDLREGMARLVDPDALVPAHFVLLPALPMSLSGKVDKLLPAPPAHADYEPPSGEL  
30 ERELAHIWQSVLHLDVRGRHDSFFDLGGHSLLAMQVLGRIESSLGIRTTLRTLFEHPTLHQLA  
DRLSSGAASTTAAATVPASEIAPSLGRAPAD  
EPYPLSYEQERLWVLEQLLPGGTAYNVVQAVRLRNLVDVDALSSALAALVRRHWSLRTVFVAS  
35 PTPQKICEPEAAPAEVVDLRGTPPDEAEAAARAWASREQATGFDLARGPVFRARLFRLDHDDVC  
VLVLSTHHIVTDAWSFQPLVRDLAELYRRARGGGPADMPPELPLQYVDFAVWQRRHLAGKRLAD  
KLAHWATLRLGLPVLELQTDPRPPVQTFRGAERVLPDLARLVAQLDELARSRGATRFMVLLA  
40 ALGVLLRRSSGQDDLAIGTAVANRPRPELEPLVGFFVNTIVMRDLGGDPTFEELLSRARKVA  
25 LEAFEHQDAPFEKVVEAVNPRRDLRSPLFQVMLVVQNAPTEALELGEVRIEPLDLPVEATRF  
DLRFSVEPRGGRDVISLQYNVDLFDATIDRMLATMQSVLSRATQDPAQVRALSVAPEDRER  
ALVAWNTAVATPDHLRLEEPFFERAVEQPDACAVVDAERRLTYGELARRAEAIAAAASRSGA  
45 TANALVAVVMEKGWEQVAAVLGVLRAGAAAYLPDLPRLPEERLRHLLHAEVRLVLTQSAVDGT  
IAWPAGIERLAVDADERWREQPVARRPPGGSTDDLAYVIYTSGSTGLPKGVMIDHRGAVNTVL  
30 DINRRFDVGPEDRVLALSSLSFDLSVYDVFGTLAAGGAVVIPDRTRASDPGHWRELVERERV

5

226

10

15

VWNSVPALMEMLMDASPGAGDPALSSLRLVMSGDWIPLKLPDRIRAACRAPRVVS LGGATEA  
SIWSIAHPIADVDPAWRSIPYGRPLANQHTYVLDEGLEPCPIGVPGEIHIGGIGVALGYWRDE  
ARTRERFLKHPTTGERLYRTGDLGRYFADGTIELLGRTDHQVKIRGFRIELGEIEAALAQHPS  
VEQAVVAAKTDPGSEKRLVAYVVGADGDGAALRDFVRKKLPEYMIPAEVVVLPALPLSANGKV  
5 DRAALPDPAAVAPRAAAVAPRTATERLIASVLAEVLQVEAVGVTDNLFELGFTSLLLVRQRL  
15 LAERIAARAPDEGAAAQAVSLTDLFQYPTIEQLAQRLDAATVKAEPADVGAQRAEARRDARRR  
RGRG

20

25

30

SEQ ID No 110 (Contig 12 >ORF1)  
10 PPAVRRYVADRRPEQLPALAPEEREAAAARLSALGAAPPQVRRRGLTRAPLSYGQSRIYFLEQ  
LSPGKPLFNVPGAVRLRGPVDVARLSAAPGEIVRRHDLRTSIANVDGELLQIAQPHAGFALD  
VVTSTPEEAAELDRRLRAEAWRPFAIGAPPLLRATLFRLAEDHVLLVTMHHVVSDDWSLGI  
LRELLALYAGRSLPPRLQVSDFAAWQREMVESGALDQRAYWRERLRGLSRASISAGGGAEA  
PSHDPGATEEIALSPDKAAALEALARREGATLFMVLLALLDLVIHARSGALDIAGTPIANR  
15 NRPELEDVVGLLTNTLVIRVDLARAGAFRDVLARARVQALDAFANQDIPFDVVTQDLKQERDH  
AQHPLFRVWLALQNAPKPALEVRGLRVEPLPLRPELVHFEVALLWPADDGSLVGHFEFRRDR  
VDEGARKEIAAAFTHLVDIAVIRPDAPVSTLVEGARAEAAQAALGEAFARAATARLGQLRR  
30 RSAGDRTPRE

35

40

45

20 SEQ ID No 111 (Contig 12 >ORF2)  
MSIHIEEEGRADAAKAPPFQYLQALHSALAHENDPVKRKQIEAGMVFKWLREEPLPFLSQLRR  
EKPISIPAITLVTRYNDVVEVLNANDVFSVDNIAPKLVENVGQNILAMNDSPKYEHEKSLLR  
LAFPRADLPYRQIVVDEANRLAKVGVDKPFDLTGDYALRVPAGAMARYLGVGGEI PTEKVVA  
40 WTHALFNEIFLNPTNDPTAVAAARAARQEALPMIDAIVAARKKQLAKSPPEQPSVLDRYLM  
25 QSVPETYESDEGIRDVILGLLMGCVDLSGGAIVNALVELMKRPRVLRDALNVVNVEDDAAITG  
YVLEALRFRPPSTGVTSLCVRDVTVGRGTRHEEKVPAGALVMACSASAMHDHEHIDAPDQFRP  
GRLP SRNYLFWESGIHTCHGKYVAILHISLAIKQLLRAGVPSAIDPMPRVHGYPAFVRVRLAA  
45 AEG

50

55

5

227

SEQ ID No 112 (Contig 12 &gt;ORF3)

10

MREPSSTPNWRNFGSNLPAGSDSVPPGEGFPPIKKILALNLGKWKDTAGLQIAQALHLFEYGYK  
RYREGKFVLRATSDLGGLGAIFESIDNWESFDQFEEFFK?WTFIRKPLVATRWAEAEFGQRQL  
VGINPAHIRRATPADLADFVSGAEPKPIAIADGRTLEEVREGGQLYFLDYRIFKDIVDITDVQE  
5 ELGKYPLAPTCLHQTAAAGELLPVAIRLVHSRPGKGAHPDKIFTSPSGPSDDWLTAKIAVASAD  
15 AIYQGQVTHLLYAHLIVEPFAVSTYRNLPATHPLHQLLRPHFFNTLAINELARRRFLGRGRFF  
DITSSVATMGSEFELLTRAYTGKGIKGYGGKPPWRFYESALPRDLSARDVRLVGYHYRDDALLH  
WDAIQEYVGQVLKIAYPPTPGSLSSDASLQRWIHELVSQGGMSLLPPERADQLEKLTSLDD  
LIAIVTNIIFTATAYHAAVNFGQTDYYTWIPNAQFATYRSYGDVLNGSEKRQFKPLERLPGRA  
20 QSIRQMVLRSLSMGPPILTSESLMTMKCLLQDPAAKQAFARYRERLAHIEREITERNRAREQP  
10 YLYLLPSMVPQSVAI

25

SEQ ID No 113 (Contig 12 &gt;ORF4)

15

VSSSRSTGRVPRDRASPAGSCAPALVGPPLSYASVMPPLDLHVALFGASGAGKTVLLAAFYR  
AQTPQSFQOEYAYKIQAVNKAQGNQLGRFYRLEEGRFPDGSTRFDEYEFDFFPDLPEPAVR  
IHWYDYPGRWWEDEPVDADEREAMRQGLIRLGMSQVGILLADGAKYRAEGTGYIRWLFHFAD  
30 ECDRLRRASAATGDEVSFPREWILALSKADLCPDYSARDFEREVCRDADDQLAKLCSVLRAE  
HAFGHRFMLLSVAAPAGAQVDPRTSLGVRTLAPAILVSTVEGAVREAQAARKEKSAGETFFQ  
GLRDLVQFVDSLDDFLPKRYQIVSKILRFISIKDFATTRLDRLKKMREDAIRKGDFTFAVLTA  
20 MVAALRDDEGARAYHQNQ

35

SEQ ID No 114 (Contig 12 &gt;ORF5)

40

MPAPAPLVETSRLLRTRGEHWYEFICVPEIPALPAWLSTLEAMLADADAGAGELRYGLLEI  
DDRGQRAPRAYPYVAVRFLDPARRDWTGRQVQHFAAWFPPVPEAVAELPEAVPADWHLRVLD  
25 GLAGTYGSGEVFGLEPATIRAWKRSHDESRAARAMAIVKATPPVSLGGGEAAPSRWTRVPTLK  
KKPPEPPAAAGLLSVGAVPSGQGRRFGCFAIGAMMLAAFCRLMLACGVRLGGA

45

SEQ ID No 115 (Contig 12 &gt;ORF6)

30

VRFRSSLGPLLLAALGAALTVSAAFRSAEASVFDSASRPEDADGHVRIPVCIDPTSSAEQRV  
DGAAGGLIHAPNPSLADVITRVRTALQGSWERWSSVRFTGWESCDSSLPATRMTYGVGRIHPD

50

55

5

10

APNQSDSIGVYNKGGSVQFKPWGADFNRCIKYNWQTARVEYSFDCVEQYAIHEMGHAIGFMHE  
WHHPLVPSACSQREPLPASDVASGWPPSSRRYIVVNPFGFYDYSIMTYWSGCSDQDGVRFSGSET  
LDAVDIQAVATVYPPVGGAPDVCNPGWFAGKRWFCAAQPTVSVGNSCSSGWVECLPHCNPRPF  
QGEWWTCTPNPYAVTGQSCSARWELCGD

5

15

SEQ ID No 116 (Contig 12 >ORF7)  
VGESQGALVGGNALSTNALNLNALNLNALNLNLSGLSARNLAAIQDPGPGSALARDFLRY  
AASCALSSTASFDFSWTDSNGKRHDERYPGLLGVA PAWASGPLDDAGQRIVSSCVAARVNYQ  
VPVLLSARSLRDPLKTLSSSQELIDYPDVEGAFWGNLFAAQPYINACYN SATVDNSRAYQRDC  
AAGHVTSGGQIVECGLIRIAGSCDRVCQKLNAGAGQYYPSCVDRPGQSTATTKDVITTALP

20

10

25

SEQ ID No 117 (Contig 12 >ORF8)  
VLAHCERGGLTARAASLLARGAELAAARRAYLDAEGCYGRVEALLGALLPEERRARGLARFR  
LGRHTEALADLAAAREAAAAAASEAGAEIELLLDEAMILDWTGEYRAARERVAAAERLAGRVAS  
PLLGARLLLGVRSLHRADREDEAAAVLTRAAAQAARLGDEGHETHIIALLLLGFILASLGRV  
EEAARDLDAVILSCEERSDLMHLGAALNNRGLARALQGDRA GMIADFERTIALGRELGQPAFE  
LVGRYNLAEYLYLMDDLAAARPHARAVQAIAPRCGDRHAPVVVTL LIARLRLYQGDEAGARRI  
ALRLRAARDDAGCEALKPSEDVLCAMIELATRDDRAAWAALEERSARCSVGQERIEVLEARA  
LAALRRGRRADARAQLERALAAASTIPTVMGGRLRRWYAE LTRATESDAPDIDLAAAEATFTG  
ARAREKVEY

30

20

35

40

25

45

30

SEQ ID No 118 (Contig 12 >ORF9)  
QAYPDLWAERGRQELWLRQLPPRACAQLAREALGDAADGALIDRLVTQSEGQPPFFLEELIRAT  
AEGRGDALPETVVAMVQVRLEALAPPARRILRAASVLGEVFWRGAVAHLLGGDEAAPLA EHLS  
ALVAGELCVRHREGRFPGEEEYSFRQALLREGAYAQLTKDDRALGHR LAADWLEAAGEADPLV  
LAHCERGGLTARAASLLARGAELAAARRAYLDAEGCYGRVEALLGALLPEERRARGLARFRL  
GRHTEALADLAAAREAAAAAASEAGAEIELLLDEAMILDWTGEYRAARERVAAAERLAGRVASP  
LLGARLLLGVRSLHRADREDEAAAVLTRAAAQAARLGDEGHETHIIALLLLGFILASLGRVE  
EAARDLDAVILSCEERSDLMHLGAALNNRGLARALQGDRA GMIADFERTIALGRELGQPAFEL  
VGRYNLAEYLYLMDDLAAARPHARAVQAIAPRCGDRHAPVVVTL LIARLRLYQGDEAGARRIA

50

55

5

10

LRLRAARDDAGCEALKPSEDVLCAMIELATRDDDRAAWAALEERSARCSVGQERIEVLEARAL  
AALRRGRRADARAQLERALAAASTIPTVMGGRLLRRWYAELTRATESDAPDIDLAAA EATFTGA  
RAREKVEY

15

- 5 14. DNA sequence according to any of claims 1 to 5 wherein the  
DNA is selected from the group consisting of

(a) the following DNA sequences:

20

10 Seq ID No 119 (>Contig17)

25

TTACGTTACTCATCCTATCTCGGCACCCTGTGTCGGTGATGTGCTCGCC  
TCGAGCGCGAGCGGGACGACGTCGGCGCCGCGCTCGGTGAGCGCCGCCGC  
GAGGGCGCTCGCGAGATCGCTGGCGACGCCGGCCGGGGCCACGACGAGCC  
ACGTCCCCGCGACGTCGCCCGGTGACGCGGCGCTCACGGGTCTCCATTCTG  
15 ACGCGGTAGCGCCACGCGCCACGGTGCTCTGCTCTCGGCGGCTCCGCCG  
CCACGCCGACAGGGCCGGCATGAGGCTCTCGAGGGCCGAGCGCCGCCCGC  
30 TGTCGGCGACGTGGAGCGGTCCGAGAGCGCCGACGTCGCCGCGCTCG  
ATGGCTCGCCAGAACCGGTCTCCTCGGCGGACGCTCCCGGCGCCGCGTC  
CTCATCGTCCGACGCGTCGCCTGCGTCGAGCCAGAACCGCTCGCGCTGGA  
20 ACGCGTACGTCGGCAACGTCACGCGGCGCGCCCGAGCGGAGCGAAGAAC  
35 GCACCCAGTCGATGGCGTGCCCGCGCGCGTGGAGCTGCCTGCCGAGAG  
GAGGAAGCGCTCGAGGTGCCTTCGTGCGGCGGAGCGAGGACACCACGG  
TCGCATCGCCGTCGATCGACGAGAGCGTCTCGTCGAGCGCGACGGTGAGC  
40 ACGGGGTGAGGGCTGACCTCGACGAAGAAGCGGTGGCCGTCGTCGAGCAG  
25 GCGCGCGGTGGCGTGCTCGAAGCGGACGGTGTGGCGCAGGTTTCGGTACC  
AGTGGGCGGCGCCGAGGGCCTCGCCATCAAGCCTCTCGCCCGTCACCGCG  
GASTAGAGCGGCACGGTCGCCGGGCGCGGCGGATGCCGTCGAGCGCCTC  
45 CAGCATCGTCCGCTCGATGGCCTCCACGTGGGCGGAGTGGGAGGCGTACT  
CGACGCGGACCTTGCGGGCGAACAGCTGCGCCCGCTCAGCTCTGCGACG  
30 AGTTCGTCGATAGCGCCGGGTCTCCGAGACGAGGGCCGCGTGAGGGCT

50

55

5

230

10

15

20

25

30

35

40

45

50

55

GTTGATCGCCGCTATCGCCAGGCGTTCGCCCAAGGGCGCAAGGCGCGCCT  
CGAGCTCGGCGGTGGTGAGCTCGACGGCGGACATGGCGCCGCTCCCGCG  
AGCTTCGTAATGGCGCGCGAGCGGAGCGCGACACCCTGGCGGCGTCTTC  
TAGCGAGAGCGCGCCCGCGACGTACGCGGCCGCGATCTCGCCCTGGCTGT  
5 GGGCGACGACCGCGTTCGGGCGTGACTCCGGCGGCGCGCCAGGTGGCGGCG  
15 AGGGCGATCATGACGGCGAACAGCACGGGCTGCACCACGTGACGCGCTC  
GAGCATGGGCGCGGCTGCGCTTCGTGCGCGCCGAGCACGGCGAGGAGCG  
ACCACTCGACGTGCGGCGCCAGGGCGCGCTCGCACGCCTCGATCTCGGCC  
CGAAAGGCGGGCGAGGAGCGAGCAGAGCGCGGCCATCGATGGCCACTG  
20 CGAGCCCTGGCCGGGAAGACGAAGGCGACCTTGCCCGGCGGGAGCGCCT  
CGCCCGCGACCGTTCTGCCCCGCGCGCCCTCGGCGAGCGCCGCGAGC  
GCCGAGAGCAGCGCGGCGCGATCGTCTGCCACGACGGCGGCGCGACGCTC  
25 GAAATGCGACCGCGTGGTTCGCGAGCGACGCCGCGACGTGACGAGGGCGA  
CGTCCTCGTGCTCGGCGAGGTGCGCGTGAGCTTGCCCGCCTGAGCGCGG  
15 AGCGCCGCGTCGCTCTTCGCCGAGAGGAGCACCGGCACCGGCGGCGCGAA  
GGGCGCGCGGGCGGGCTCCCCGGCTTGGTCGTGCGCGGCCGCGCGCG  
30 GCGCTTCTTCGAGGACCAGTGCAGGTTGGTGCCGAGATCCCGAACGAC  
GACACCGCCGCGCGCCGAGGAGACCCGCTGGCTTCCACGGTACCTCCTC  
GGTCAAGAGGCGGATCGCGCCGCGACGACCAATCGATGTGCTGCGACGGGC  
20 TCGCGGCGTGAGCGTCTTCGGGAGGACGCCGCTCTGCAGCGCGAGCACC  
ATCTTGATGACCGCGCCGATCCCCGCGGCGGCTGCGTGTGCCGAGGTT  
CGACTTTAGGCTCCCGAGCCACAGCGGGCGCTCCTTCGCGTGCGCCGCGC  
CGTACGTGCGGAAGAGCGCGCGCGCTCGATGGGATCGCCGAGCGTCGTG  
40 CCGGTTCCGTGCGCCTCGACGGCGTCGACGTCCGCGGGGCGAGCCCCGC  
25 GCTCGCGAGCGCGTCCCGGATCACGCGCTCTTGCGCGGGGCCGTTTCGGCG  
CCGTGAGCCCTTGGCTCTTGCCGTCCTGGTTGACGGCCGATCCGCGCACG  
ATCGCGAGCACGGGGTGCCCGTCTTCCGGGCGTCCGACAGGCGCTCGAG  
45 GAGCACTATCCAGCGCCTTCCGACCAGCCCGCGCGGTTTCGCGTGCGACG  
AGAACGACTTGACACGCCCCGTCCGGCGCGCCCGCGTGTGCGCGCTGAAC  
30 TCGCCGAAGATCCCGGGGGTCCGCATCACGGTCACGCCGCCGCGGAGCGC

5

231

10

GAGCGAGCACTCGCCTCGACGGATGGCGTGGCAGGCGAGGTGGAGCGCGA  
CGAGCGACGAGCTGCACGCCGTGTCGACGCT

Seq ID No 120 (>Contig18)

15

20

25

30

35

40

45

50

55

5 TTTTAGGANCCCCGACGTGCACGATCGGCTCGCCAACTCGTGGCGCGCC  
GGGACTATTTTACCAGCTCGCGTTGCGCGCCGCGGGACCTACGTGCGG  
GGCCTCGTCCGCGCCCCGACGACGGCGCGCGCCCCCGCGTTCGCGCC  
GCGTGGGGCGGCGCTCGTCACGGGCGGGACCGGGCGCTCGGGGCGCACG  
TTGCCCGTTGGTTGCGCGGATCGGCGCCGAGCACATCGTGCTCGCGAGC  
10 CGCCGCGGAGCGCGGCCCGGCGCGGCCGCGCTCGCCGAGGAGCTTTC  
GGTGCTCGGCGCGCGCTGACGCTGCTTGGCTGCGACGTCCCGATCGTG  
AGGCGGTGCGGGGCTCGTGCGCAACGTCAAGGCCGCGGAGCGACGGTG  
CGCGCCGTGTTCCACGCGGGCGGTGCGATGCACGAGGCGCCGGTCGCCGC  
CATGCGTGTGAGGAGCTCGCCGACGCGATCGCCGTGAAGGCCCGCGGCG  
15 CGCAGCACCTCCAAGACGTCTTCGCGCAGCGCCCGCTCAACGCGTTTGTC  
CTCTTCTCGTCAGAAACGGGTGTGTGGGCGGTGGCCGGAAGGCGCGTA  
CGCCGCGGCAACGCGTTCTCTGACGCGCTCGCCGAGGCGCGTCGCGCGG  
ACGGCCTCGCGGCGACCTCGATCGCGTGGGGCGCGTGGGCGGGCGCGGA  
ATGCTCGCGACCGACGCCGAGCGGCGCTTGAAGCATCGCGGCGTCGCGCC  
20 GATGGATCCGGAGCTCGCCGTGCGGGCCCTCGCGACGCGCTCGATCACG  
CCGAGACGTGCCTCGCCGTGCTGACGTGACTGGGCGCGCTTCGCCCCG  
TCGTTGCGCTCGGCGCGTCTCTGCCCCGCTCCTCGACGAGCTCGCGGAGGC  
GCGATCGGCGCTCGACGCGCTGCGCGAGCCACCGACGACGCGCGCACGG  
40 CCGCCGGTCCCAGCCCCGAAGCACGCTGAGGACCACGCTCGCGGCGCTC  
25 CCGGAGGGCGAGCGCCACCGCCACCTCCTCGCGCTCGTGCGGACGGAGAC  
GGCGGCGGTGCTCGGGCACGCGGACGCGTGGCGCGTCGAGCCGAACCGCG  
GGTTCCTTTGACCTCGGGCTCGACTCGCTCATGTCCGTGAGCTCCGCAGG  
CGCGTCCAGCGCGGACCGGCATCAAGCTCCCGGCGACGCTCGCGTTCGA  
CCACCCGACGCCGAGCGCGCTCGCGAGCAAGGTGCTCGCCGCGATCGTCC  
30 TCCACGACGCGACCCCGCGCGCCTCGCCCCCGCGGAGCTCGAGCGCCTC



5

232

10

5

15

20

10

25

15

30

35

20

40

25

45

30

50

55

GAGGGGATGCTCTCGGCGATCTACGCGGACGAAGCGCTCCGCGACGACCT  
CACGGCGCGCCTCCGCGCCTTCCTGGACAAGCGCGCGGTCCGCACCGAAC  
GCCCCGACGACGCCGCGTTTCGCCGAGAAGCTCGGCTCCGCGAGCGCCGAC  
GAACTCATTGCGCTGATCGATCAGAAGCTCGGAGATCGCATCGATGTCGA  
CCGTTACTAACGACACGCTCACGGAGTACTTGGCGCGCCTCACTCAAGAG  
CTCCACAGGAGCGAGACGCGCCTGCGTGCGACGGAAGAGAGGCGACATGA  
GCCGATCGCCATCGTCGGCCTCGGGCTCCCCCTCCGGGGCGGGATCCACG  
ACCGCGACACGCTCTGGACGTTCTCGAGGAGGGCCGCGACGCCATCGCG  
CCGATCCTCGCGAGCCGCTGGAACGCGGACGCGACGTACGACCTCGATCC  
GGACGCCGTCGGCAAGAGCTACGTGCGCGACGCCGCCATGCTCGATCGCG  
TCGACCTTTTCGACGCCGATTCTTCGGGATCAGCCCGCGCGAGGCGAAG  
TACGTCGACCCGACGACCGCCTCTTGCTCGAGACGTCGTGGCAAGCGCT  
CGAGGACGCGGGGATTGTGCCGGCGTCGCTGCGAGACTCGAAGACCGGCG  
TCTTCGTCGGCACGGGCGCGAGCGACTACGCGTTCCTCCAGAGCGATCGC  
GACGCCTCGGAGGCGTACGCGTTTCATGGGGATGATCTCGTCGTTTCGCGGC  
GGGCCGCTTCGCGTTCACGCTCGGGCTCCAAGGCCCGCGCTATCGATCG  
ACACGGCGTGCTCTTCGTCGCTCGTCGCGCTCCACCTCGCGTGCCAGTCG  
CTGCGTCAAGGCGAGTGCGACCTCGCGCTCGTCGCGGTGTGCAGGTCAT  
GTCGTCGCCGAGGTGTTTCGTGCTGCTCTCGCGCACGCGCGCTCGCGA  
GCGACGGGCGATCGAAGACGTTCTCGGCGAACGCCGACGGCTATGGCCGC  
GGCGAAGGCGTCGTCTCCTGGCCGTCGAGCGCCTCCGCGACGCGCGCGC  
GAAAGGGCGCCCGATCCTCGCGGTGATCCGCGGACGCGCGGTGAACCACG  
ACGGCACGTCGAGCGGGATCACGGTCCCGAACGGGCCCGCGCAGCAGAAG  
GTGCTCCGCGCCGCGCTCGACGACGCGCGGCTTGTCGCCCGCGACGTCGA  
CGTCGTCGAGTGCCACGGCACGGGGACCTCCATCGGCGATCCCATCGAAG  
TGAACGCGCTCGCCGCCGTCTACGGGAGGGGCGCCCAAGGACCGCCCG  
CTGTTCTCGGGCGCGCTGAAGACCAACATCGGGCACCTCGAGTTCGCGTC  
GGGCCTCGCCGGCGTCGCGAAGATGGTCGCCTCCATGCGCCACGCGACCC  
TCCCCGCGACGCTGCACACGAGCCCGCTCAACCCGCTCGTCGACTGGGAC  
GCGCTCCCCGTCGCGCTCGTCGACGCCGCGCGCCCGTGGACGCGCCGCGA

5

233

10

15

20

25

30

35

40

CGACGGCGCCCCCGGCGCGCCGGCGTCACGGCGATCGTCGAGGAGGCGC  
CCGCCGAGCCCGAGCCACGACGCCCCGACGCCGCGCCCGCGCTTCCGGCC  
GTGCCCCTTCTCCTCTCGGGCAAGACCGACGAGGCGCTGCGCGCGCAGGC  
AGCGCGCCTCCACGCGCACTCGCGGGGCGCCCCGACGCGCGGCTCGTCG  
5 ACATCGCCGCGTCGCTCGCGACGACGCGCACGCACTTCGATCGACGCGCG  
GCCGTCGTCGCGGCGGATCGCGACGAGCTCCTCGGCGCGCTCGACGCGCT  
15 CGCGCGCGGCGAGGCAGGCCCGGGGTCGGTCGTCGCGAGCGCGATCCCCG  
CCGGCAGGGTCGTGTTCTGTGTTCCCCGGCCAAGGCTCGCAGTGGGTCTGGG  
ATGGCGCGCGCTCCTCGCGTCGTCGGTGGTCTTCCGCGACGAGATCGC  
10 GGCTGCGAGCGCGCGCTCGCGCCGACGTCGCCTGGTCGCTCGGCGCCG  
TTCTCCGGGGCGACGCGGACGAGGCGACGCTCCTCGGCCGCGTCGACGTC  
GTGCAGCCGGTCCTCTTCGCCGTCATGGTCGCCCTCGCCGCGCTCTGGCG  
25 CTCGATCGGCGTCACGCCCCGACGCCGTCGTCGGGCACAGCCAAGGCGAGA  
TCGCCGCCGCCTACGTCGCCGGCGCCCTCTCGCTCGAAGACGCCGCCAAG  
15 GTCGTCGCGCTGCGCGCACGAGCGCTCACGAAGATCGCGGGGCGCGGGG  
GATGGCCGCCGTCGAGCTCGGCGCACGCGACACCGAGGCGCGCCTCGCGC  
30 CGTTCGGCGACGCCATCGCGATCGCGGCGATCAACAGCCCGCGGCCACG  
CTCGTCGCGGGCGACACGGACGCGATCGACGCGCTCGTCCGCGACCTCGA  
GGCCGCGCAGATCTTCGCGCGGAAGGTGCGTGTCTGACTACGCGTCGCACT  
20 CGGCGCACGTCGAGGCGATCGAGCGCGAGCTCCTCGCGGATCTCGCGGG  
ATCGAACCGCGCGCGGGCGCTGTGCCGCTTTACTCCGCGGTGACGGGCGC  
GAAGCTCGACGGGAACCGCCTCGACCCCGCGCATTGGTTCCGGAACCTGC  
GCTCGACAAAAAATTTGAGGACGCCACGCGCGCGCTCCACGACGACGGC  
40 CGCCGGGTATCCTCATNATCNGGGCGTNCAGAGGAGTCGGTATTNCCCC  
25 CCCCCGCCTTNCCCG,

or their complementary strands,

45

50

55

5

10

(b) DNA-sequences which hybridise under stringent conditions to regions of DNA-sequences according to (a) encoding proteins or to fragments of said DNA-sequences,

15

(c) DNA-sequences which hybridise to the DNA-sequences according to (a) and (b) because of a degeneration of the genetic code,

20

(d) allele variations and mutants resulting by substitution, insertion or deletion of nucleotides or inversion of nucleotide segments of DNA-sequences according to (a) to (c), wherein the variations and mutants offer isofunctional expression products.

25

15. Peptide encoded by a DNA sequence according to claim 14 selected from the group consisting of

30

Seq ID No 121

>Contig17\_001 828 amino acids MW=86259 D pI=5.60 numambig=1

35

20 MTVMATPGIFGEFSAQHAGAPDGRCKSFSSHANGAGWSEGAGIVLLERLSDARKNGHPVL  
AIVRGS AVNQDGKSQGLTAPNGPAQERVIRDALASAGLAPADVDAVEAHGTGTTLGDP  
ARALFATYGAAHAKERPLWLGLSKSNLGHQTQAAAGIGGVIMVLALQSGVLPRTLHAASP  
SQHIDWSSGAIRLLTTEVPWKPGGSPRRAAVSSFGISGTNAHVLEEAPRAAGDDQAGE  
PARAPFAPPVPVLLSAKSDAALRAQAGXLHAHLAEHEDVALVDVAASLATTRSHFERRAA  
40 VVADDRAALLSALAALAEGRAGAGTVAGEALPPGKVAFVFPQGGSQWPSMARALLASSPA  
25 FRAEIEACERALAPHVDWSSLAVLGGDEAHAAPMLERVDVVPVLFVMIALATWRAAG  
VTPDAVVGHSSQGEIAAAYVAGALSLEDAARVVALRSRAITKLARGAMSAVELTTAELEA  
RLAPLGERLAIAAINSPHAALVSGDPGAIDELVAELSGAQLFARKVRVEYASHSAHVEAI  
45 ERTMLEALDGIAPRPATVPLYSAVTGERLDGEALGAHWYRNLRHRTVRFEHATRALDDG  
HRFFVEVSPHPVLTVALDETLLSSIDGATVVSSLRRDEGDLERFLLSAGELHARGHAIDW  
30 GAFFAPLGARRVTLPTYAFQRRERFWLDAGDASDDEDAAPGASAEETAFWRAIERGDVAAL

50

55

5

235

10

SDALHVADSGRRSALESIMPALSAWRRSRREQSTVGAWRYRVEWRPVSAASRGDVAGTWL  
VVAPAGVASDLASALAAALTEGADVVPALALEASDITDTGCRDRMSNVX

Seq ID No 122

15

5 >Contig18\_002 502 amino acids MW=53019 D pI=6.83 numambig=1  
FRXPDVHDLRLANLVARRDYFYQLALRAAGTYVRGLVRAPHD GARPPAFAPRGAALVTGGT  
GALGAHVVARW FARIGA EHVILASRRGAAAPGAAALAEELSVLGARVTLVACDVPDREAVA  
GLVRNVKAGGATVRAVFHAGGAMHEAPVAAMRVEELADAIIVKARGAQHLQDVFQQRPLN  
20 AFVLFSSSETGVWGGGRQGAYAAANAFDALAEARRADGLAATSIAWGAWAGGGMLATDAE  
10 RRLKHRGVAPMDPELAVAALAHALDHAETCLAVADVWARFAPS FASARPRPLLELAEAE  
RSALDALREPPDDARTAAGPEPASTLRTTLAALPEGERHRHLLALVRTETA AVLGHADAS  
RVEPNRGFFDLGLDSLMSVELRRRVQRATGIKLPATLAFDHPTPSALASKVLAAIVLHDA  
25 TPRASPAEELERLEGMLSAIYADEALRDDLTARLRAFLDKRAVRTERPDDAFAEKLGSA  
SADELIRLIDQKLGDRIIDVDY\*

25

15

Seq ID No 123

30

>Contig18\_010 840 amino acids MW=88062 D pI=5.74 numambig=6  
MSTVTNDTLTEYLRLRLTQELHRSETRLRATEERRHEPIAIVGLGLPFRGGIHDRDTLWTF  
LEEGRDAIAPILASRW NADATYDLDFDAVGKSYVRDAAMLDRVDFDADFFGISPREAKY  
35 VDPQHRLLLET SWQALE DAGIVPASLRDSKTGVFVGTGASDYAFLQSDRDASEAYAFMG  
20 ISSFAAGRLAFTLGLQGPALSIDTACSSSLVALHLACQSLRQGECDLALVAGVQVMSSPE  
VFLLSRTRALASDGRSKTFSANADGYGRGEGVVVLAVERLRDARAKGRPILAVIRGSAV  
NHDGTSSGITVPNGPAQQKVLRAALDDARLVPADVDVVECHGTGTSIGDPIEVNALAAVY  
40 GEGRPKDRPLFLGALKTNIGHLEFASGLAGVAKMVASMRHATLPATLHTSPLNPLVDWDA  
25 LPVRVVD AARPWTRRDDGAPRRAGVTAIVEEAPAEPEPTTPDAAPALPAVPVLLSGKTDE  
ALRAQAARLHAHLAGRPDARLVDIAASLATTRTHFDRAAVVAADRDELLGALDARGE  
AGPGSVVASAIPAGRVVFVFPQGQSQWVG MARALLASSVFRDEIAACERALAPHVAVSL  
45 GAVLRGDGDEATLLGRVDVVQPVLFAVMVALAALWRSIGVTPDAVVGHSSQGEIAAAYVAG  
ALSLEDAAKVVALRARALTKIAGRGAMAAVELGARDTEARLAPFGDAIAIAAINS PRATL  
30 VAGDTDAIDALVRDLEAAQIFARKVRVDYASHSAHVEAIERELLADLAGIEPRAGAVPLY

50

55

5

236

SAVTGAKLDGNRLDPAHWFRNLRSTKNFEDATRALHDDGRRVSSXSXAXRGVGIXPPRLX  
X

10

16. Recombinant expression vector which comprises a DNA-  
5 sequence according to any of claims 1 to 10, 12 and 14.

15

17. Procaryotic or eucaryotic cell which has been transfected  
or transformed with a DNA-sequence according to any of claims 1  
20 to 10, 12 and 14 or with a recombinant expression vector ac-  
10 cording to claim 16.

20

18. Cell according to claim 17, wherein the cell is derived  
25 from myxobacteria.

25

19. Cell according to claim 17, wherein the cell is derived  
from a Sorangium strain.

30

20. Cell according to claim 17, wherein the cell is derived  
from Sorangium cellulosum.

35

21. Cell according to claim 17, wherein the cell is derived  
from a Streptomyces strain.

22. Cell according to claim 17, wherein the cell is derived  
25 from Escherichia coli.

40

23. Process for an enzymatic biosynthesis, mutasynthesis or  
45 partial synthesis of polyketide or heteropolyketide compounds,  
wherein a cell according to any of claims 17 to 22 is culti-

50

55

5

237

10

vated in a suitable culture medium and the polyketide or heteropolyketide compound is isolated from the medium.

15

24. Process according to claim 23, wherein the polyketide or heteropolyketide compound is an epothilone.

20

25

30

35

40

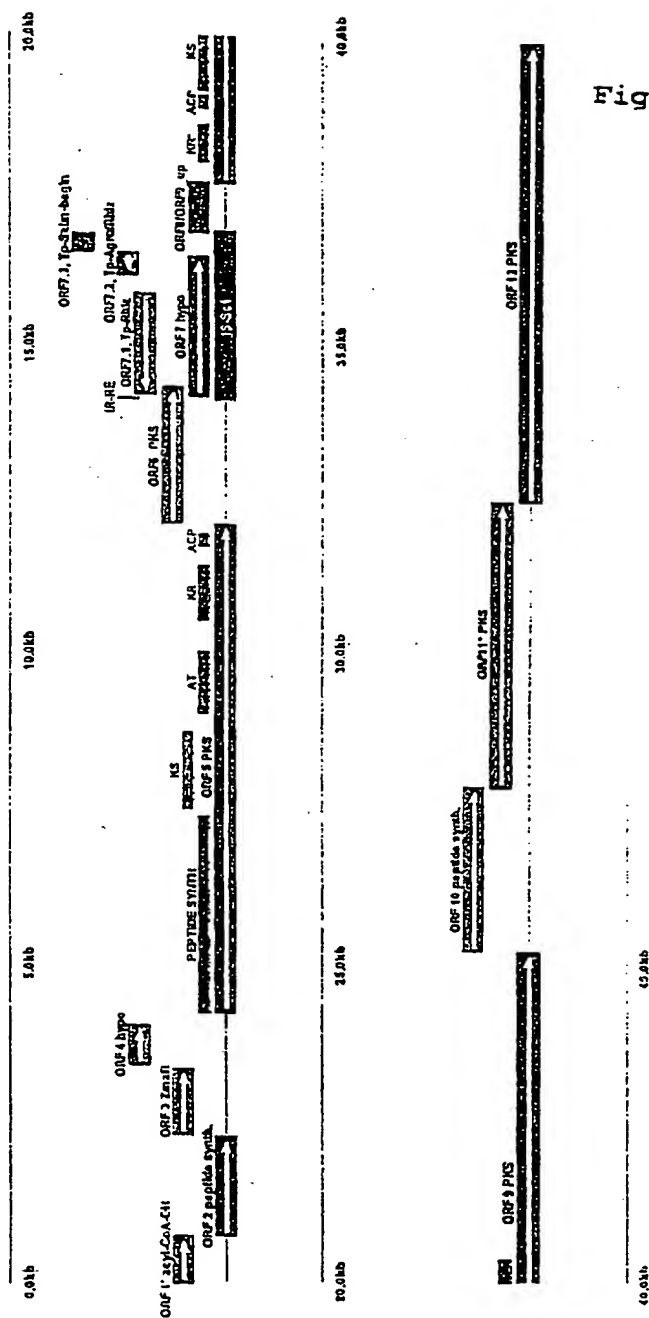
45

50

55



Open reading frames found on pEPOcos6 region.  
pEPOcos6\_annotated  
47713 bp





(19) World Intellectual Property Organization  
International Bureau(43) International Publication Date  
20 April 2000 (20.04.2000)

PCT

(10) International Publication Number  
WO 00/22139 A3(51) International Patent Classification<sup>7</sup>: C12N 15/52,  
9/00, 15/63, 5/10, C12P 17/06, C07K 14/195, C12P 17/18

(21) International Application Number: PCT/US99/23535

(22) International Filing Date: 11 October 1999 (11.10.1999)

(25) Filing Language: English

(26) Publication Language: English

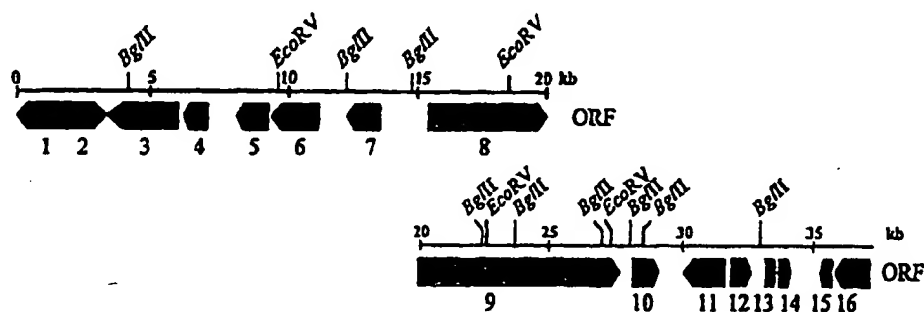
(30) Priority Data:  
198 46 493.2 9 October 1998 (09.10.1998) DE(71) Applicants (for all designated States except US):  
GESELLSCHAFT FÜR BIOTECHNOLOGISCHE  
FORSCHUNG MBH (GBF) [DE/DE]; Mascheroder  
Weg 1, D-38124 Braunschweig (DE). BRISTOL-MY-  
ERS SQUIBB COMPANY [US/US]; P.O. Box 4000,  
Princeton, NJ 08543-4000 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): BEYER, Stefan  
[DE/DE]; Mascheroder Weg 1, D-38124 Braunschweig  
(DE). BLOECKER, Helmut [DE/DE]; Mascheroder  
Weg 1, D-38124 Braunschweig (DE). BRANDT, Petra[DE/DE]; Mascheroder Weg 1, D-38124 Braunschweig  
(DE). CINO, Paul, M. [US/US]; 4 Crest Drive, Bound  
Brook, NJ 08805 (US). DOUGHERTY, Brian, A.  
[US/US]; 10 Rosemary Lane, Killingworth, CT 06419  
(US). GOLDBERG, Steven, L. [US/US]; 25 Compton  
Court, Basking Ridge, NJ 07920 (US). HOFLE, Gerhard  
[DE/DE]; Mascheroder Weg 1, D-38124 Braunschweig  
(DE). MUELLER, Rolf-Joachim [DE/DE]; Mascheroder  
Weg 1, D-38124 Braunschweig (DE). REICHENBACH,  
Hans [DE/DE]; Mascheroder Weg 1, D-38124 Braun-  
schweig (DE).(74) Agents: SANTUCCI, Ronald, R. et al.; Kane, Dalsimer,  
Sullivan and Levy, LLP 20th floor, 711 Third Avenue, New  
York, NY 10017 (US).(81) Designated States (national): AE, AL, AM, AT, AU, AZ,  
BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK,  
DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL,  
IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU,  
LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT,  
RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA,  
UG, US, UZ, VN, YU, ZA, ZW.(84) Designated States (regional): ARIPO patent (GH, GM,  
KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent

[Continued on next page]

(54) Title: DNA SEQUENCES FOR ENZYMATIC SYNTHESIS OF POLYKETIDE OR HETEROPOLYKETIDE COMPOUNDS



(57) Abstract: The invention consists of: (1) cloned *Sorangium cellulosum* polyketide synthase (PKS) biosynthetic cluster DNA; and (2) the nucleotide sequence and predicted protein coding sequences of the cloned DNA. The invention can be used for, but not limited to: (a) increasing yields of PKS product in *Sorangium cellulosum* (e.g., by amplification or genetic modification of the epothilone gene cluster or its component parts); (b) increasing yields of polyketide product in a heterologous system by transfer of the epothilone gene cluster or its component parts, which may be followed by amplification or genetic modification of the PKS gene cluster or its component parts; (c) modification of the polyketide product chemical structure in either *Sorangium cellulosum* or a heterologous host (e.g., by genetic modification of the epothilone gene cluster or its component parts; and (d) for the detection of genes and gene products involved in making polyketides or related molecules in other organisms (e.g., by hybridization or complementation assays). DNA sequence and analysis is presented for the following cosmids and plasmids: A2 cosmid; the pEPOcos6 region (overlapping of pEPOcos6 and pEPOcos7); pEPOcos8 cosmid; A5 cosmid; Sau4 (10 kb plasmid).



(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**(15) Information about Correction:**

**Previous Correction:**

see PCT Gazette No. 36/2000 of 8 September 2000, Section II

**Published:**

— *With international search report.*

**(88) Date of publication of the international search report:**

18 January 2001

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 99/23535

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/52 C12N9/00 C12N15/63 C12N5/10 C12P17/06  
 C07K14/195 C12P17/18

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X  A	<p>US 5 716 849 A (SCHUPP THOMAS ET AL)          10 February 1998 (1998-02-10)</p> <p>abstract          column 1, line 15 - line 23          column 4, line 8 - line 33          column 4, line 66 - column 5, line 40          example 1</p> <p style="text-align: center;">--- -/--</p>	<p>1,3-5,          16,17,          22,23          18-21</p>

☒ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.

## \* Special categories of cited documents:

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

- \*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- \*Z\* document member of the same patent family

Date of the actual completion of the international search

18 August 2000

Date of mailing of the international search report

23. 10. 00

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
 NL - 2280 HV Rijswijk  
 Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
 Fax: (+31-70) 340-3016

Authorized officer

CEDER O.

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 99/23535

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SCHUPP T ET AL: "SORANGIUM CELLULOSUM (MYXOBACTERIUM) GENE CLUSTER FOR THE BIOSYNTHESIS OF THE MACROLIDE ANTIBIOTIC SORAPHEN A: CLONING, CHARACTERIZATION, AND HOMOLOGY TO POLYKETIDE SYNTHASE GENES FROM ACTINOMYCETES" JOURNAL OF BACTERIOLOGY, vol. 177, no. 13, July 1995 (1995-07), pages 3673-3679, XP000893003 WASHINGTON, DC ISSN: 0021-9193 the whole document ---	1,3-5, 16,17, 22,23
X	WO 93 13663 A (ABBOTT LAB) 22 July 1993 (1993-07-22) abstract ---	1,16,17, 21,23
A	WO 98 22461 A (BIOTECHNOLOG FORSCHUNG GMBH ;GERTH KLAUS (DE); HOEFLE GERHARD (DE)) 28 May 1998 (1998-05-28) the whole document ---	2,24
P,X	BEYER S ET AL: "METABOLIC DIVERSITY IN MYXOBACTERIA: IDENTIFICATION OF THE MYXALAMID AND THE STIGMATELLIN BIOSYNTHETIC GENE CLUSTER OF STIGMATELLA AURANTIACA SG A15 AND A COMBINED POLYKETIDE(POLY)PEPTIDE GENE CLUSTER FROM THE EPOTHILONE PRODUCING STRAIN SORANGIUM CELLULOSUM SO CE90" BIOCHIMICA ET BIOPHYSICA ACTA. GENE STRUCTURE AND EXPRESSION, vol. 1445, no. 2, 14 May 1999 (1999-05-14), pages 185-195, XP000915670 AMSTERDAM ISSN: 0167-4781 the whole document ---	1-5
E	WO 99 66028 A (NOVARTIS ERFINDE VERWALT GMBH ;NOVARTIS AG (CH); SCHUPP THOMAS (CH)) 23 December 1999 (1999-12-23) the whole document ---	1-6,8, 16-24
E	WO 00 31247 A (KOSAN BIOSCIENCES INC) 2 June 2000 (2000-06-02) the whole document -----	1-6,8, 16-24

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US 99/23535

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
  
2. ☐ Claims Nos.:  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
  
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
  
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
  
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
  
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:  
1-5 and 16-24 partly and 6,8 completely

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

1. Claims: 1-5 and 16-24 partly and 6,8 completely

Nucleic acid and polypeptide sequences involved in polyketide synthesis and related to cosmid A2 and their uses, where the sequences are seq id nos 1,3-34.

2. Claims: 1-5 and 16-24 partly and 7,9 completely

Nucleic acid and polypeptide sequences involved in polyketide synthesis and related to the pEP0cos6 region and their uses, where the sequences are seq id nos 2,35-70.

3. Claims: 1-5 and 16-24 partly and 10,11 completely

Nucleic acid and polypeptide sequences involved in polyketide synthesis and related to the pEP0cos8 region and their uses, where the sequences are seq id nos 71-88.

4. Claims: 1-5 and 16-24 partly and 12,13 completely

Nucleic acid and polypeptide sequences involved in polyketide synthesis and related to the cosmid A5 and their uses, where the sequences are seq id nos 89-118.

5. Claims: 1-5 and 16-24 partly and 14,15 completely

Nucleic acid and polypeptide sequences involved in polyketide synthesis and related to plasmid Sau4 and their uses, where the sequences are seq id nos 119-123.

## INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 99/23535

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5716849 A	10-02-1998	US 5639949 A	17-06-1997
		AU 695364 B	13-08-1998
		AU 2417895 A	04-01-1996
		BG 101107 A	28-11-1997
		BR 9507948 A	18-11-1997
		CA 2192366 A	14-12-1995
		CN 1152941 A	25-06-1997
		CZ 9603601 A	14-01-1998
		EP 0759078 A	26-02-1997
		HU 76534 A	29-09-1997
		WO 9533818 A	14-12-1995
		JP 10501125 T	03-02-1998
		NZ 285166 A	27-04-1998
		PL 317998 A	12-05-1997
		SK 156996 A	06-08-1997
		US 6117670 A	12-09-2000
		US 5670350 A	23-09-1997
		US 5643774 A	01-07-1997
		US 5723759 A	03-03-1998
		US 5679560 A	21-10-1997
		US 5698425 A	16-12-1997
		US 5710031 A	20-01-1998
		US 5686282 A	11-11-1997
		US 5686283 A	11-11-1997
		US 5817502 A	06-10-1998
		ZA 9504686 A	08-12-1995
-----	-----	-----	-----
WO 9313663 A	22-07-1993	CA 2100791 A	18-07-1993
		AU 665526 B	11-01-1996
		AU 1245092 A	03-08-1993
		EP 0626806 A	07-12-1994
-----	-----	-----	-----
WO 9822461 A	28-05-1998	AU 5483798 A	10-06-1998
		BR 9713363 A	25-01-2000
		CN 1237970 A	08-12-1999
		CZ 9901750 A	15-09-1999
		EP 0941227 A	15-09-1999
		NO 992338 A	14-05-1999
		PL 333435 A	06-12-1999
		ZA 9710384 A	18-05-1999
-----	-----	-----	-----
WO 9966028 A	23-12-1999	AU 4611699 A	05-01-2000
		US 6121029 A	19-09-2000
-----	-----	-----	-----
WO 0031247 -A	02-06-2000	AU 1737700 A	13-06-2000
-----	-----	-----	-----